



# Nouvelles approches combinant protéomique, immuno-enrichissement et bioinformatique pour la détection de microorganismes

Emie Durighello

## ► To cite this version:

Emie Durighello. Nouvelles approches combinant protéomique, immuno-enrichissement et bioinformatique pour la détection de microorganismes. Biochimie, Biologie Moléculaire. Université Montpellier I, 2014. Français. NNT : 2014MON13514 . tel-01132152

**HAL Id: tel-01132152**

**<https://theses.hal.science/tel-01132152>**

Submitted on 16 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Pour obtenir le grade de  
**Docteur**

Délivré par **UNIVERSITE de MONTPELLIER I**

Préparée au sein de l'école doctorale Sciences Chimiques et  
Biologiques pour la Santé (ED168)  
Et de l'unité de recherche Laboratoire de Biologie des  
Systèmes Perturbés (CEA, DSV)

Spécialité : **Biochimie et Biologie Moléculaire**

Présentée par **Emie DURIGHELLO**

**Nouvelles approches combinant  
protéomique, immuno-enrichissement et  
bioinformatique pour la détection de  
microorganismes**

**Soutenue le 16 décembre 2014 devant le jury composé de**

M. Jean ARMENGAUD, Directeur de Recherche, CEA Marcoule

Directeur de thèse

Mme Virginie BRUN, Chercheur, CEA Grenoble

Rapporteur

M. Patrick FOURQUET, Ingénieur de Recherche, INSERM Marseille

Rapporteur

Mme Catherine DUPORT, Professeur, INRA Avignon

Examineur

Mme Christine ENJALBAL, Professeur, CNRS Montpellier 2

Président du Jury

M. Laurent BELLANGER, Chercheur, CEA Marcoule





*La vie est un ballet ; on ne le danse qu'une fois.*

Proverbe Malinké



# Remerciements



# Remerciements

Je tiens à remercier l'ensemble des membres du jury qui ont accepté de juger mon travail. Je remercie Dr. Brun et Dr. Fourquet d'avoir accepté d'être rapporteurs ainsi que Dr. Duport et Dr. Enjalbal d'être examinateurs et Dr Bellanger de prendre part au jury. A tous, merci d'avoir pris de votre temps pour évaluer mon travail.

Merci à Jean Armengaud de m'avoir accueillie au sein de son laboratoire et d'avoir guidé mon travail de thèse tout au long de ses trois années. Merci à Laurent Bellanger de m'avoir donné la chance de profiter des ressources et des connaissances de son laboratoire. Merci également à Eric Ezan et à Pierre Chagvardieff qui se sont succédé à la tête du service et qui ont permis que ma thèse se déroule dans de bonnes conditions.

Je voudrais aussi remercier les personnes travaillant à Bertin Technologies, à Bruker et au Robert Koch Institut avec qui j'ai pu collaborer grâce au projet GEFREASE.

Merci à Cindy Morris, Catherine Duport et Béatrice Alpha-Bazin d'avoir accepté de faire partie de mon comité de thèse et de m'avoir donné des conseils à différentes étapes de mon projet de thèse. Merci également à Christine Almunia, Béatrice Alpha-Bazin et Laetitia Pinto pour leur relecture attentive et minutieuse.

Je remercie tout le personnel que j'ai côtoyé au sein du SBTN et plus particulièrement Alain Lorphelin, Marie-Anne Roncato, Jean-Charles Gaillard et Philippe Guérin qui ont parfois été mes mains ou ma tête et sans qui ce travail n'aurait pas pu être réalisé. Merci également à Christine Almunia, Béatrice Alpha-Bazin, Valérie Aloin, Béatrice Alonso, Béatrice Baus, Carole Béraud, Yves Brignon, Anne-Hélène Davin, Nicole Desmoulière, Claude Dubois, Océane Ducos-Anglade, Bernard Fernandez, Fabrice Gallais, Christian Godon, Céline Guigue, Thi-ngoc Suong Huynh, Sandra Laloi, Emeline Landas, Sarah Le-Hir, Charles Marchetti, Sabrina Masmejean, Mickael Mege, Guylaine Miotello, Nicole Sage, Yasmina Tayeb, Ali Younes, qu'il était toujours agréable de croiser aux détours d'une paillasse, d'un couloir ou d'un café.

Merci à tous les membres du planning gâteaux, merveilleuse idée pour finir la semaine sur une note sucrée et garder un poids de forme toute l'année.

Mille mercis et bien plus encore à mes *poupines* du travail : Céline Bland, Carine Darolles-Dolgopyatioff et Laetitia Pinto qui ont mis et continuent de mettre des paillettes (et des potins !) dans ma vie, et à François Allain, co-bureau, collègue, colocataire, copain, qui me fait voir la vie en rose.

Bon courage aux futurs docteurs du SBTN : Cédric Pisani, Jean-Paul Madeira, Thi-ngoc Suong Huynh et à Judith Trapp et Alain Dedieu pour qui l'échéance est toute proche. Toutes mes pensées aussi à mes camarades de l'ESPCI qui sont dans le même bateau ou sortis depuis

peu : Louise Bonnemay, Julie Bouchon, Antoine Cagniant, Marie Delafontaine, Johanna Goldman, Laurent Lermusiaux, Amir Nahas, Marine Verhulssel.

Un immense merci à toute l'équipe du SMBP et en particulier à Giovanni Chiappetta, Marie Couzinie, Emmanuelle Demey, Iman Haddad, Segal Ndiaye, Yann Verdier et Joëlle Vinh grâce à qui j'ai fait mes premiers pas en protéomique.

Merci à mes amis de lycée, à mes copains d'école et aux filles de la GR (ça gère) qui sont un peu trop loin géographiquement à mon goût, mais toujours prêts à faire la fête quand je passe par la capitale.

Et pour finir je ne remercierai jamais assez mes parents et toute ma famille de m'épauler et de toujours être là pour moi dans les bons comme dans les mauvais jours.

# Résumé





## Nouvelles approches combinant protéomique, immuno-enrichissement et bioinformatique pour la détection de microorganismes

### Résumé

Identifier rapidement des microorganismes pathogènes dans des échantillons environnementaux est un enjeu majeur dans le domaine de la biodéfense. Dans ce contexte, la spectrométrie de masse MALDI-TOF peut offrir une réponse simple, rapide et peu coûteuse. L'enjeu de la thèse, dans le cadre du projet ANR franco-allemand GEFREASE, a été de développer des méthodes permettant l'identification des microorganismes pathogènes et notamment de mettre en place des approches ciblées pour la préparation d'échantillon à l'aide d'anticorps en amont de la spectrométrie de masse. Dans un premier temps, l'étude du protéome de la bactérie modèle, *Francisella tularensis* subsp. *holarctica* LVS, responsable de la tularémie, a permis d'identifier les protéines et les peptides les plus abondants donnant un signal intense par spectrométrie de masse. Ensuite l'étude protéogénomique de douze protéines cibles a permis de choisir trois biomarqueurs dont le profil des masses par spectrométrie de masse de type MALDI-TOF (approche top-down) est spécifique de l'espèce et de la sous-espèce des bactéries du genre *Francisella*. Par cette méthode la virulence d'une souche est donc rapidement déterminée puisqu'elle est dépendante de la sous-espèce à laquelle la bactérie appartient. Ce test mis au point présente l'avantage d'être simple et rapide. Dans un deuxième temps, la mise au point d'un protocole d'enrichissement de la bactérie modèle par immunocapture magnétique a permis de montrer qu'il est possible de concentrer des bactéries grâce à des billes magnétiques couplées à des anticorps dirigés contre la bactérie entière. Cette approche a été expérimentée dans le cas de mélanges de bactéries où la bactérie modèle était largement minoritaire et dans des échantillons de matrices alimentaires diverses telles que de l'eau minérale ou du lait. La méthodologie a été validée sur un agent de classe 3, *Francisella tularensis* subsp. *tularensis*.

Mots clés : Protéomique, MALDI-TOF, biomarqueurs, immunocapture, microbiologie, *Francisella tularensis*



## New approaches for microorganisms detection combining proteomics, immuno-enrichment and bioinformatics

### Abstract

The rapid identification of pathogenic microorganisms in environmental samples is a major issue in the biodefense field. MALDI-TOF mass spectrometry can offer a fast, straightforward and inexpensive answer. In the framework of the Franco-German ANR project GEFREASE, the purpose of the thesis was to develop methodologies allowing identification of pathogenic microorganisms and particularly to set up targeted approaches using antibodies for sample preparation before mass spectrometry. First of all, the proteome study of *Francisella tularensis* subsp. *holarctica* LVS, responsible for tularemia, allowed us to identify the most abundant proteins and peptides for which the most intense signals are observed when using mass spectrometry. The proteogenomic study of twelve of these proteins enable us to choose three biomarkers for which the masses monitored by MALDI-TOF mass spectrometry (top-down approach) allow deciphering the *Francisella* species and subspecies. The interest of this work is being able to conclude on a strain virulence based on the knowledge of the subspecies it belongs. The finalized test is easy and fast. Secondly, the development of a magnetic immunocapture of *Francisella tularensis* subsp. *holarctica* LVS allowed us to show that it is possible to concentrate bacteria using magnetic beads coupled to antibodies raised against the entire bacterium. This approach has been experimented in the case of bacterial mixtures where the model bacterium was largely in minority and for samples containing various food matrices such as mineral water or milk. The methodology has been validated on a class 3 agent, *Francisella tularensis* subsp. *tularensis*.

Keywords: Proteomics, MALDI-TOF, biomarkers, immunocapture, microbiology, *Francisella tularensis*



# Table des matières



## Table des matières

<b>Remerciements</b>	7
<b>Résumé</b>	11
<b>Abstract</b>	13
<b>Table des illustrations</b>	22
<b>Table des tableaux</b>	24
<b>Liste des abréviations</b>	25
<b>Introduction bibliographique</b>	31
1 Les organismes pathogènes et les agents de la menace	31
1.1 Un peu d'histoire en matière d'armes biologiques et de bioterrorisme	32
1.2 Les microorganismes listés comme relevant du risque bioterroriste	33
1.2.1 Liste des agents	33
1.2.2 Exemples d'évènements impliquant ces microorganismes et répartition géographique naturelle	34
1.3 <i>Francisella tularensis</i> : agent de la tularémie	35
1.3.1 Description de la bactérie	36
1.3.2 Mode de culture	36
1.3.3 La maladie	36
1.3.4 Répartition géographique naturelle	36
1.3.5 Virulence	36
1.4 Les méthodes de détection et d'identification existantes	37
1.4.1 Culture et tests phénotypiques	37
1.4.2 Méthodes génomiques	38
1.4.3 Méthodes immunologiques	39
1.4.4 Méthodes de spectrométrie de masse	40
2 Les apports de la protéomique et des nouveaux spectromètres de masse	40
2.1 Le « boom de la protéomique » en lien avec l'évolution de la spectrométrie de masse	41
2.2 La spectrométrie de masse	42
2.2.1 Les sources d'ionisation	42
2.2.2 Les analyseurs	44
2.2.3 Le principe de fragmentation	46
2.2.4 Les méthodes de fragmentation	48
2.3 Shotgun et quantification, les nouvelles approches	49



2.3.1	La technique shotgun .....	49
2.3.2	Méthodes de quantification .....	51
3	Interprétation des spectres MS et MS/MS .....	53
3.1	Attribution de séquences peptidiques aux spectres MS/MS .....	53
3.1.1	Attribution à partir d'une base de données – Le logiciel Mascot .....	53
3.1.2	Séquençage de novo .....	55
3.1.3	Recherche d'étiquette .....	55
3.2	Identification de microorganismes .....	56
3.2.1	Un peu d'histoire .....	56
3.2.2	Exemples de méthodes d'identification de microorganismes .....	57
3.3	Systèmes intégrés d'identification de microorganismes .....	59
3.3.1	Biotyper (Bruker) .....	60
3.3.2	Andromas (Andromas SAS) .....	60
3.3.3	VITEK® MS (bioMérieux) .....	61
3.4	Publications des résultats d'identification .....	61
4	Méthodes d'enrichissement de bactérie pour une identification par spectrométrie de masse MALDI-TOF. ....	62
4.1	L'immunocapture magnétique .....	62
4.2	Les nanoparticules (NPs) .....	63
4.3	Les surfaces fonctionnalisées – SELDI .....	63
5	Objectifs du travail de thèse .....	65
5.1	Le projet Franco-Allemand GEFREASE .....	65
5.2	Les objectifs du projet de thèse entrepris .....	66
	<b>Résultats</b> .....	67
	<b>1<sup>ère</sup> partie : Identification de <i>Francisella tularensis</i> jusqu'au niveau sous-espèce par la détection de biomarqueurs spécifiques</b> .....	69
1	Contexte et stratégie proposée .....	69
2	Caractérisation des protéines les plus abondantes par analyse protéomique .....	70
2.1	Détection MALDI-TOF après inactivation au TFA .....	70
2.2	Gel 1D, digestion in-gel et analyse nanoLC-MS/MS .....	73
2.3	Masses des protéines entières et modifications post-traductionnelles .....	75
2.4	Identification des protéines majoritaires .....	75
3	Identification de biomarqueurs potentiels par approche protéogénomique .....	78
4	Choix des biomarqueurs .....	84
5	Validation des biomarqueurs .....	85
6	Conclusion .....	88

<b>2<sup>ème</sup> partie : Immunocapture de <i>Francisella tularensis</i> couplée à une détection MALDI-TOF</b>	
91	
1 Contexte et stratégie proposée .....	91
2 Méthodologie choisie.....	91
3 Mise en œuvre des anticorps.....	92
3.1 Fabrication des anticorps monoclonaux .....	92
3.2 Cible des anticorps .....	93
3.2.1 Données structurales sur le LPS .....	93
3.2.2 Données fonctionnelles sur le LPS .....	94
4 Les billes magnétiques .....	95
5 Méthodes analytiques pour l'évaluation des résultats.....	96
5.1 Mesure d'absorbance.....	96
5.2 Test ELISA .....	96
5.3 Spectrométrie de masse MALDI-TOF .....	97
6 Optimisation du protocole d'immunocapture .....	97
6.1 Choix de l'anticorps .....	97
6.2 Couplage des anticorps .....	98
6.2.1 Test ELISA .....	99
6.2.2 Analyse par spectrométrie de masse MALDI-TOF .....	100
6.3 Cible de l'immunocapture .....	102
6.4 Inactivation.....	103
6.5 Lavages .....	106
6.6 Elution .....	108
6.6.1 Tampon.....	108
6.6.2 Volume .....	108
6.7 Récapitulatif du protocole de couplage et d'immunocapture.....	109
6.7.1 Protocole de couplage des billes magnétiques.....	109
6.7.2 Protocole d'immunocapture des bactéries actives.....	109
7 Optimisation et validation de l'approche MS .....	110
7.1 Analyse MALDI-TOF .....	110
7.1.1 Matrice .....	110
7.1.2 Les différentes conditions .....	110
7.1.3 Calibration .....	114
7.2 Traitement des données MALDI-TOF .....	114
7.2.1 Traitement des spectres bruts .....	114
7.2.2 Algorithme de comparaison des profils protéiques.....	115

8	Banc d'essai .....	118
8.1	Test de matrices .....	118
8.1.1	Essais en matrice PBS .....	118
8.1.2	Essais en matrice lait UHT demi-écrémé.....	121
8.1.3	Essais en matrice eau d'Evian .....	124
8.2	Mélanges hétérogènes de bactéries .....	127
8.3	Echantillons dilués.....	131
9	Remise en culture de bactéries après immunocapture .....	132
10	Conclusions.....	135
10.1	Les inconvénients de cette méthode .....	135
10.1.1	Limite de détection .....	135
10.1.2	Manque de discrimination niveau sous-espèce.....	135
10.2	Les avantages de cette méthode .....	135
10.2.1	Rapidité .....	135
10.2.2	Robustesse .....	136
10.2.3	Automatisation.....	136
10.2.4	Culture .....	136
	<b>Discussion.....</b>	<b>139</b>
1	Apports et complémentarité des deux méthodes développées. ....	139
2	Application à d'autres microorganismes de la menace .....	139
3	Détection de nouveaux biomarqueurs à partir de l'exoprotéome.....	141
4	Détection environnementale .....	142
5	Le diagnostic médical : vers plus de sensibilité ? .....	142
6	Nouveautés instrumentales .....	143
	<b>Matériels et méthodes.....</b>	<b>147</b>
1	Recherche de biomarqueurs .....	147
1.1	Culture des bactéries.....	147
1.2	Méthodes d'inactivation. ....	147
1.3	Validation de l'inactivation.....	148
1.4	Analyse en spectrométrie de masse .....	148
1.4.1	MALDI-TOF .....	148
1.4.2	NanoLC-MS/MS .....	148
1.4.3	Traitement des données MS .....	150
1.4.4	Analyse des séquences protéiques et nucléotidiques .....	150
2	Immunocapture magnétique .....	151
2.1	Anticorps .....	151

2.1.1	Développement des anticorps .....	151
2.1.2	Test ELISA .....	151
2.2	Préparation des billes magnétiques.....	152
2.3	Les bactéries.....	153
2.4	Culture des bactéries.....	153
2.5	Préparation des bactéries .....	153
2.6	Comptage des bactéries.....	154
2.7	Immunocapture magnétique des bactéries.....	154
2.8	Inactivation des bactéries à la chaleur.....	154
2.9	Remise en culture des bactéries .....	154
2.10	Analyse en spectrométrie de masse .....	155
2.10.1	MALDI-TOF .....	155
2.10.2	Validation des résultats.....	155
	<b>Références bibliographiques</b> .....	159
	<b>Annexes</b> .....	173
1	Table des annexes .....	173
2	Participation à l'écriture de revue et d'articles.....	185
2.1	Articles.....	185
2.2	Revue.....	186
3	Liste des communications réalisées durant cette thèse :.....	213
3.1	Présentations orales :.....	213
3.2	Posters :.....	213

# Table des illustrations

<b>Figure 1</b> : Taxonomie de <i>Francisella</i> .....	35
<b>Figure 2</b> : Plage d'identification selon les méthodes génomiques..	38
<b>Figure 3</b> : Schéma général d'un spectromètre de masse.....	42
<b>Figure 4</b> : Mode linéaire et mode réflectron du tube de vol .....	45
<b>Figure 5</b> : Nomenclature peptidique.....	46
<b>Figure 6</b> : Théorie du proton mobile.....	47
<b>Figure 7</b> : Structure des ions fragments d, v et w. ....	48
<b>Figure 8</b> : Identification de protéines par MS/MS. ....	50
<b>Figure 9</b> : Fonctionnement général d'un algorithme d'identification de protéines .....	54
<b>Figure 10</b> : Identification par recherche de signature d'étiquette. ....	55
<b>Figure 11</b> : Comparaison de profils protéiques pour l'identification de microorganismes .....	60
<b>Figure 12</b> : Interaction des nanoparticules avec la bactérie selon leur toxicité..	63
<b>Figure 13</b> : Stratégie d'identification des biomarqueurs .....	70
<b>Figure 14</b> : Spectre MALDI-TOF de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS avec calibration interne .....	72
<b>Figure 15</b> : Gel d'électrophorèse 1D du protéome de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS inactivée au TFA.....	73
<b>Figure 16</b> : Protéines identifiées par nanoLC-MS/MS et caractéristiques fonctionnelles.....	74
<b>Figure 17</b> : Spectre MALDI-TOF de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS avec calibration interne.....	76
<b>Figure 18</b> : Démarche protéogénomique.....	80
<b>Figure 19</b> : Evolution des profils MALDI-TOF de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS au cours de 11 jours de culture à 37°C. ....	85
<b>Figure 20</b> : Spectre MALDI-TOF de <i>Francisella philomiragia</i> ATCC 25015.....	86
<b>Figure 21</b> : Spectre MALDI-TOF de <i>Francisella tularensis</i> subsp. <i>tularensis</i> CCUG 2112.....	87
<b>Figure 22</b> : Schéma du LPS .....	94
<b>Figure 23</b> : Schéma réactionnel entre un sulphonyl ester et le groupement amine d'un anticorps.....	95
<b>Figure 24</b> : Schéma général du protocole d'immunocapture.....	97
<b>Figure 25</b> : Récapitulatif des différentes étapes de l'immunocapture .....	99
<b>Figure 26</b> : Spectre MALDI-TOF de l'éluat des billes <i>Activ-MasterBeads</i> couplées à 50 µg d'Ac/mg de billes.....	101
<b>Figure 27</b> : Spectres MALDI-TOF des éluats des billes <i>Dynabeads MyOne® Tosylactivated™</i> .....	102
<b>Figure 28</b> : Spectres MALDI-TOF des éluats d'immunocapture .....	103
<b>Figure 29</b> : Spectres MALDI-TOF, éluats des billes couplées aux anticorps FT100R de 10 <sup>9</sup> CFU/ml à 10 <sup>6</sup> CFU/ml. ....	104

<b>Figure 30</b> : Spectres MALDI-TOF, éluats des billes couplées aux anticorps TMPP de $10^9$ CFU/ml à $10^6$ CFU/ml. ....	105
<b>Figure 31</b> : Spectre MALDI-TOF, éluat de l'échantillon sans bactéries, chauffé 1h à 90°C....	105
<b>Figure 32</b> : Spectre MALDI-TOF, éluat de l'échantillon sans bactéries et sans chauffage.....	105
<b>Figure 33</b> : Spectres MALDI-TOF des immunocaptures selon les différentes conditions de lavages.....	107
<b>Figure 34</b> : Spectres MALDI-TOF des bactéries selon différentes conditions.....	112
<b>Figure 35</b> : Spectres MALDI-TOF des bactéries à deux concentrations dans de l'eau avec les matrices SA et HCCA.....	113
<b>Figure 37</b> : Scores moyens sur 3 tirs sur un dépôt d'éluat.....	120
<b>Figure 37</b> : Scores moyens des immunocapture dans du lait .....	122
<b>Figure 38</b> : Spectre MALDI-TOF de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS à $10^9$ CFU/ml dans du lait U.H.T demi-écrémé.....	123
<b>Figure 39</b> : Scores moyens des immunocapture dans de l'Evian.....	126
<b>Figure 40</b> : Spectres MALDI-TOF des solutions de bactéries avant immunocapture .....	129
<b>Figure 41</b> : Scores moyens des témoins positifs et négatifs. ....	130
<b>Figure 42</b> : Scores moyens de la solution à $10^6$ CFU/ml et à $10^7$ CFU/ml après pré-concentration au Cell-Trap et immunocapture. ....	131
<b>Figure 43</b> : Photos des boîtes de gélose ensemencées avec 1 µl de billes magnétiques.....	133
<b>Figure 44</b> : Spectre MALDI-TOF d'une remise en culture d'une immunocapture de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS inactivée au TFA.....	134

# Table des tableaux

<b>Tableau 1</b> : Matrices chromophores et leurs utilisations. ....	43
<b>Tableau 2</b> : Mode de calcul des masses moyennes des protéines.. ....	75
<b>Tableau 3</b> : Liste des 12 protéines identifiées. ....	77
<b>Tableau 4</b> : Liste des souches de <i>Francisella</i> dont le génome est partiellement ou complètement séquencé.....	78
<b>Tableau 5</b> : Observation des différences de séquences et de masses pour la protéine HU forme B, selon les espèces de <i>Francisella</i> . ....	79
<b>Tableau 6</b> : Liste des masses de 12 biomarqueurs pour les 24 souches de <i>Francisella</i> .....	82
<b>Tableau 7</b> : Combinaisons de masses spécifiques aux espèces et aux sous-espèces de <i>Francisella</i> .....	84
<b>Tableau 8</b> : Rendement de couplage des différents types de billes.....	98
<b>Tableau 9</b> : Résultats des tests ELISA pour déterminer la concentration de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS dans les solutions aux différentes étapes de l'immunocapture .....	100
<b>Tableau 10</b> : Répartition des lavages en H <sub>2</sub> O et en PBS selon les tubes. ....	106
<b>Tableau 11</b> : Optimisation de la quantité de billes et du volume d'élution. ....	108
<b>Tableau 12</b> : Conditions de tampons, de ZipTip et de matrices utilisées .....	111
<b>Tableau 13</b> : Signification des scores de l'algorithme d'identification .....	117
<b>Tableau 14</b> : Résultats de l'algorithme de reconnaissance de <i>Francisella tularensis</i> subsp. <i>holarctica</i> LVS .....	117
<b>Tableau 15</b> : Récapitulatif des échantillons préparés en PBS.....	118
<b>Tableau 16</b> : Méthode de préparation des échantillons de lait. ....	121
<b>Tableau 17</b> : Composition des différents témoins d'immunocapture.....	121
<b>Tableau 18</b> : Méthode de préparation des échantillons d'eau d'Evian.....	124
<b>Tableau 19</b> : Composition des différents témoins d'immunocapture.....	124
<b>Tableau 20</b> : Méthode de préparation des mélanges bactériens.....	128

# Liste des abréviations

AA : Acide aminé  
Ac : Anticorps  
ACN : Acétonitrile  
ACP : Analyse en Composante Principale  
ADN : Acide désoxyribonucléique  
AFLP : Polymorphisme de longueur des fragments amplifiés, Amplification Fragment Length Polymorphism  
ANR : Agence Nationale de la Recherche  
ANSM : Agence Nationale de Sécurité du Médicament et des produits de santé  
ARN : Acide ribonucléique  
BLAST : Basic Local Alignement Search Tool  
BHI : Infusion cerveau cœur, Brain Heart Infusion  
bp : Paire de base  
BTS : Bacterial Test Standard  
CEA : Commissariat à l'Energie Atomique et aux Energies Alternatives  
CDC : Centre de Contrôle et de prévention des Maladies, Centers for Disease Control and prevention  
CFU : Unité formant des colonies, Colony-Forming Unit  
CID : Dissociation induite par collision, Collision Induced Dissociation  
DO : Densité Optique  
DTT : Dithiothréitol  
ECD : Dissociation par capture d'électron, Electron Capture Dissociation  
ELISA : Dosage d'immunoabsorption par enzyme liée, Enzyme-Linked Immunosorbent Assay  
ESI : Ionisation par électroébuliseur, Electrospray Ionization  
ETD : Dissociation par transfert d'électron, Electron-Transfer Dissociation  
eV : Electronvolt  
FDA : Agence des produits alimentaires et médicamenteux, Food and Drug Administration  
FDR : Taux de faux positifs, False Discovery Rate  
FT-ICR : Résonance cyclotronique d'ions à transformée de Fourier, Fourier Transform Ion Cyclotron Resonance  
GEFREASE : Equipements franco-allemands pour l'analyse et la surveillance des menaces biologiques dans l'environnement, GERman FRENch Equipment for Analysis and Surveillance of biothreats in the Environment  
GOLD : Genome Online Database  
GSB : Tampon glycérol salin  
HCCA : Acide  $\alpha$ -cyano-4-hydroxycinnamique  
HCD : Higher-energy Collisional Dissociation  
ICAT : Isotope Coded Affinity Tags



IR : Intensité Relative  
 IRMa : Interprétation des Résultats MASCOT  
 iTRAQ : isobaric Tags for Relative and Absolute Quantification  
 L2 : Laboratoire de sécurité biologique de niveau 2  
 LB : Bouillon Lysogène  
 LICB : Laboratoire d'Ingénierie Cellulaire et Biotechnologie  
 LVS : Souche vaccinale, Live Vaccine Strain  
 LPS : Lipopolysaccharide  
 MALDI : Matrix Assisted Laser desorption/Ionization  
 MLST : Typage génomique multilocus, Multilocus Sequence Typing  
 MLVA : Analyse multiple de séquences courtes répétées en tandem, Multiple Loci Variable Number of Tandem Repeats Analysis  
 MOT : Microorganismes et Toxines hautement pathogènes  
 MS : Spectrométrie de masse  
 MS/MS ou MS<sup>2</sup> : Spectrométrie de masse en tandem  
 m/z : Rapport masse sur charge  
 NCBI : National Center for Biology Information  
 NRBC : Nucléaire, Radiologique, Biologique et Chimique  
 NRBC-E : Nucléaire, Radiologique, Biologique, Chimique et Explosif  
 PBS : Tampon phosphate salin  
 PCR : Réaction en chaîne par polymérase, Polymerase Chain Reaction  
 PFGE : Electrophorèse en champ pulsé, Pulse-Field Gel Electrophoresis  
 ppm : partie par million  
 PSM : Poste de Sécurité Microbiologique  
 PTMs : modifications post-traductionnelles  
 PVX : PolyViteX  
 RFLP : Polymorphisme de longueur des fragments de restriction, Restriction Fragment Length Polymorphism  
 rpm : révolutions par minute  
 RKI : Robert Koch Institut  
 SA : Acide Sinapinique  
 SBTN : Service de Biologie et de Toxicologie Nucléaire  
 SILAC : Stable Isotope Labeled Aminoacids in Cell culture  
 SNP : Polymorphisme nucléotidique, Single-Nucleotide Polymorphism  
 Sp. : espèce  
 SRM : Selected Reaction Monitoring  
 Subsp. : Sous-espèce  
 TA : Température Ambiante  
 TFA : acide trifluoroacétique  
 TMB : 3,3',5,5'-tétraméthylbenzidine  
 TOF : Time Of Flight, Temps de vol

WHO : Organisation Mondiale de la santé, World Health Organization

XIC : Extractions des chromatogrammes d'ions, Extracted ion chromatogram



# Introduction bibliographique



# Introduction bibliographique

## 1 Les organismes pathogènes et les agents de la menace

Les agents pathogènes sont des agents biologiques responsables de maladies infectieuses. Ce sont majoritairement des bactéries et des virus mais il peut également s'agir de protozoaires ou de champignons (Pirofski and Casadevall, 2012). Le pouvoir pathogène d'un organisme est défini par sa capacité à provoquer une maladie. Il dépend de son pouvoir invasif ou de sa virulence, de son pouvoir toxique et de la résistance de l'hôte. Certains microorganismes sont inoffensifs, quoique parasites, ce sont des organismes commensaux, comme les bactéries présentes à la surface de la peau ou des muqueuses chez des sujets sains. Ces mêmes bactéries peuvent également avoir un pouvoir pathogène opportuniste, c'est-à-dire qu'elles peuvent causer des infections chez des personnes immunodéprimées. C'est notamment le cas de *Pseudomonas aeruginosa*, qui a développé une résistance aux antibiotiques et qui est responsable d'un grand nombre d'infections nosocomiales (Stover et al., 2000). Par contre, d'autres microorganismes possèdent une pathogénicité stricte qui les rend toujours pathogènes (Casadevall and Pirofski, 2000).

Le pouvoir pathogène et la virulence sont deux notions bien distinctes. En effet, un microorganisme peut être pathogène mais selon les souches être plus ou moins virulent (Casadevall and Pirofski, 2001). Prenons l'exemple de *Shigella*, l'agent responsable de la shigellose, maladie diarrhéique. Quatre espèces composent le genre des *Shigella* : *S. dysenteriae* qui est responsable des épidémies brutales et *S. flexneri*, responsable de la forme endémique. Ces deux premières espèces causent les symptômes les plus graves. *S. boydii* et *S. sonnei* provoquent des maladies moins sévères. Elles sont de ce fait moins virulentes (Niyogi, 2005).

Bactéries, virus, protozoaires et champignons sont des organismes très différents de par leur organisation cellulaire et leur taille. De ce fait, les stratégies de détection et d'identification spécifiques représentent un véritable défi biologique et technologique lorsqu'il s'agit de les identifier de manière ciblée dans une matrice complexe. De plus, parfois il ne s'agit plus d'identifier un organisme pathogène mais seulement une toxine sécrétée par cet organisme qui à elle seule est pathogène. Il n'y a donc pas nécessairement présence de bactérie viable ou de matériel génétique. C'est le cas par exemple de la toxine botulique qui est sécrétée par *Clostridium botulinum*.

Certains de ces organismes pathogènes sont considérés comme de réelles menaces dans le cadre de la lutte contre le bioterrorisme. En effet leur facilité de dispersion, de transmission d'homme à homme, leur virulence ou encore l'absence de traitement efficace font d'eux de redoutables agents de la menace.

### 1.1 Un peu d'histoire en matière d'armes biologiques et de bioterrorisme

Les premières utilisations d'armes biologiques remontent au XIV<sup>ème</sup> siècle lors du siège de Kaffa (aujourd'hui, Théodosie en Crimée). La seconde épidémie de peste sévissait alors et les Tatars utilisèrent les corps de leurs hommes infectés par la peste pour les envoyer par-dessus les remparts de la ville afin de propager l'épidémie dans la population assiégée (Derbes, 1966). Plus tard, au XVIII<sup>ème</sup> siècle les anglais donnèrent des couvertures infectées par l'agent de la variole aux Indiens dans le but de les affaiblir. D'autres attaques biologiques, notamment sur les animaux, durant la première guerre mondiale, et les dégâts provoqués par les armes chimiques conduisirent à la signature du Protocole de Genève en 1925. Ce protocole prohibe l'emploi de gaz asphyxiants, toxiques ou similaires et de moyens bactériologiques en cas de conflits. Ce traité n'empêcha pas certains états de continuer leur recherche sur les armes biologiques et leur utilisation sur l'Homme. Par exemple, il est considéré que plus de 10 000 prisonniers sont morts de ce type d'expérimentations en Manchourie, lors de l'occupation japonaise en 1937. Aux Etats-Unis plusieurs sites étaient consacrés aux recherches sur ce type d'armes, avec notamment la production de plusieurs milliers de bombes contenant des spores de *Bacillus anthracis* pendant la seconde guerre mondiale et des essais de pulvérisation en aérosol sur des volontaires de *Francisella tularensis* et *Coxiella burnetti* en 1955 (Christopher et al., 1997). En 1972 est proposé un nouveau traité, la convention sur l'interdiction des armes biologiques, interdisant cette fois-ci la fabrication et le stockage des armes bactériologiques et chimiques. Malgré la signature de 179 états, les armes biologiques furent encore utilisées. Les services secrets de l'Union soviétique commirent au moins 7 assassinats à l'aide de ricine, une toxine mortelle. La ricine était placée dans des plombs percées et scellées par de la cire prévue pour fondre à la température du corps. Les plombs étaient propulsés à l'aide d'une arme à piston maquillée en parapluie.

Quels que soient les traités signés, les possibilités d'attaques par des groupes indépendants sont les plus préoccupants. En 1984, la secte « Rajneeshee » dissémina la bactérie *Salmonella typhimurium* dans des bars à salade de l'Oregon. Ce sont 751 personnes qui furent atteintes de gastroentérites provoquées par cette salmonelle (Torok et al., 1997). Plus récemment, en 2001, ravivant au plus haut niveau médiatique la menace d'attaques biologiques, 12 courriers contenant des spores de *Bacillus anthracis* ont été envoyés par la poste américaine; 22 personnes furent touchées par la maladie du charbon et malheureusement, 5 décédèrent (Jernigan et al., 2001).

En France, cette dernière attaque déclencha la création du plan Biotox destiné à préciser l'organisation gouvernementale en cas d'exécution ou de menace d'exécution d'attaque terroriste à l'aide d'armes biologiques et à définir des mesures telles que la mise en place d'équipes spécialisées et la communication aux populations menacées ([http://www.sgdsn.gouv.fr/site\\_rubrique102.html](http://www.sgdsn.gouv.fr/site_rubrique102.html)). Depuis 2010, un plan unique regroupe les plans Piratox, Piratome et Biotox, le plan Pirate NRBC (Nucléaire, Radiologique, Chimique ou Biologique).

## 1.2 Les microorganismes listés comme relevant du risque bioterroriste

Le Centre de Contrôle et de prévention des Maladies américain (CDC, Centers for Disease Control and Prevention, <http://www.cdc.gov/>) recense tous les organismes qui apparaissent comme des agents potentiels du bioterrorisme. Ils sont classés en trois listes.

### 1.2.1 Liste des agents

La liste A comprend les agents de haute priorité qui exposent le public aux risques les plus élevés. Les critères sont les suivants : i) les agents peuvent être facilement répandus ou transmis de personne à personne ; ii) le taux de mortalité ou l'impact sur la santé publique est important ; iii) une panique générale ou une perturbation de l'ordre social peuvent être provoquées ; iv) ces agents nécessitent une action spéciale de préparation de la part de la santé publique (détection, diagnostic, décontamination, traitement).

Dans cette liste se trouvent :

- *Bacillus anthracis*, la bactérie responsable de la maladie du charbon (ou anthrax) (Spencer, 2003)
- *Clostridium botulinum*, la bactérie sécrétrice de la toxine botulique, responsable du botulisme
- *Yersinia pestis*, bactérie responsable de la peste
- *Francisella tularensis*, la bactérie responsable de la tularémie
- Smallpox, le virus responsable de la variole (Moore et al., 2006)
- Filovirus et arenavirus tels que Ebola, Marburg et Lassa, Machupo, responsables de fièvres hémorragiques (Borio et al., 2002).

La liste B se compose d'agents de moins haute importance. Ils sont : i) plus difficiles à répandre ; ii) les taux de mortalité et de morbidité sont modérés ; iii) ils demandent des capacités de diagnostic particulières et une surveillance précise des maladies.

La liste des agents concernés est plus longue et disponible sur le site internet du CDC. En voici quelques-uns : les agents menaçant la sécurité de l'eau ou la sécurité alimentaire tels que *Vibrio cholerae* ou *Escherichia coli* O157:H7 et les *Salmonella*.

La dernière liste, la liste C, contient les nouveaux agents pouvant être produits dans le but de les répandre massivement. Ils sont : i) facilement disponibles ; ii) faciles à produire et à répandre ; iii) les taux de mortalité et de morbidité sont potentiellement élevés de par le nombre de personnes susceptibles d'être contaminées, avec un fort impact sur la santé publique. Il s'agit des agents, tels que le virus de Nipah ou les hantavirus, responsables de maladies infectieuses.

L'ensemble des pathogènes sous surveillance est décrit sur le site internet du CDC et entre autres sur les sites de l'Organisation Mondiale pour la Santé (<http://www.who.int/fr/>), de l'Agence Nationale de Sécurité des Médicaments et des produits de santé en France



(<http://ansm.sante.fr/>) et de l'Agence de la Santé Publique du Canada ([www.santepublique.gc.ca](http://www.santepublique.gc.ca)).

### 1.2.2 Exemples d'évènements impliquant ces microorganismes et répartition géographique naturelle

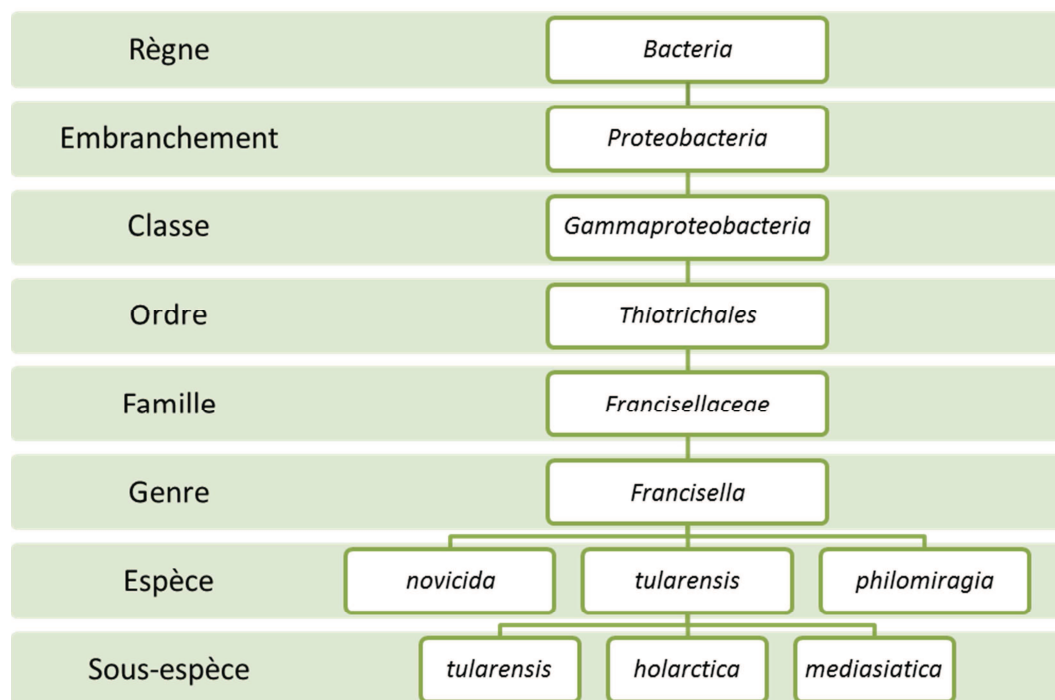
La plupart des agents des listes A et B sont présents à la surface de la terre de manière naturelle et provoquent de nombreux cas d'infections à travers le monde :

- ***Bacillus anthracis*** a vu sa plus grande épidémie au Zimbabwe avec plus de 10 000 cas d'infections entre 1979 et 1984 (Mwenye et al., 1996).
- ***Yersinia pestis*** est responsable de trois pandémies de peste au cours de ces deux derniers millénaires. La première pandémie eu lieu entre 541 et 750 après J.C. débutant à Pelusium en Egypte. La peste s'est rapidement répandue dans le bassin méditerranéen et au Moyen-Orient. La seconde pandémie qui eut lieu de 1330 à la fin du XVII<sup>ème</sup> siècle, provenant cette fois-ci d'Asie centrale, démarra réellement en Europe lors de son introduction en Sicile en 1347. C'est l'épidémie de peste noire, qui tua entre 30 à 40 % de la population de l'époque. Plusieurs épidémies eurent encore lieu de manière cyclique jusqu'à la fin du XVII<sup>ème</sup> siècle, avec des taux de mortalité plus faibles, entre 5 et 15 % (Perry and Fetherston, 1997). La troisième pandémie prit ses racines en Chine en 1855 et fut rapidement dispersée à travers le monde par le trafic maritime.
- En ce qui concerne la **variole**, le dernier cas naturellement contracté remonte à 1977 en Somalie. L'Organisation Mondiale de la Santé a déclaré que le virus avait été éradiqué de la planète en 1980 (Fenner et al., 1988).
- Pour **Ebola** les premiers cas ont été diagnostiqué en 1976 en Afrique centrale (Collectif, 1978). Oublié jusqu'en 1995, date à laquelle il a ré-émergé en République Démocratique du Congo, il est devenu l'un des agents infectieux les plus craints avec des taux de létalité pouvant aller jusqu'à 89 % (sur 143 cas en 2003 au Congo) (Feldmann et al., 2003). Cet agent est loin d'être éradiqué puisque des flambées ont lieu fréquemment, la dernière en date sévissant depuis mars 2014 en Afrique de l'Ouest est d'ampleur inégalée et n'est pour l'heure pas maîtrisée. (source : WHO, [http://www.who.int/csr/don/2014\\_04\\_02 Ebola/fr/](http://www.who.int/csr/don/2014_04_02 Ebola/fr/)).
- Le choléra, provoqué par ***Vibrio cholerae***, dont le vecteur de transmission est l'eau, est endémique de l'Asie du sud et notamment des régions du delta du Gange. A chaque pandémie cette bactérie s'est répandue un peu plus à travers le monde. Aujourd'hui la septième pandémie est mondiale après avoir démarré en Indonésie avec le biotype *V. cholerae* serogroupe O1 appelé El Tor. Le sérotype O139, avec 30 000 cas au Bangladesh en 2002, sera probablement à l'origine ce qui pourrait devenir la huitième pandémie (Sack et al., 2004).
- La tularémie provoquée par ***Francisella tularensis*** et dont les caractéristiques vont être décrites ci-dessous.

### 1.3 *Francisella tularensis* : agent de la tularémie

*Francisella tularensis*, l'agent responsable de la tularémie fait partie, selon le CDC, de la liste A des agents à surveiller. Il est considéré comme l'une des bactéries les plus virulentes parmi les bactéries pathogènes (Jones et al., 2012). En France, il fait partie de la liste des MOT (Microorganismes et Toxines hautement pathogènes) depuis l'arrêté de juin 2010. Les microorganismes et toxines faisant partie des MOT sont déterminés par l'ANSM (Agence Nationale de Sécurité du Médicament et des produits de santé). Le régime d'autorisation et les conditions pour toutes opérations de production, fabrication, transport, importation, exportation, détention, offre, cession et acquisition sont définies par des réglementations spécifiques aux MOT (<http://ansm.sante.fr/Dossiers/Micro-organismes-et-toxines-hautement-pathogenes-MOT/>).

Principalement répandu dans l'hémisphère nord, le genre *Francisella* comprend 7 espèces différentes : *Francisella tularensis*, *Francisella novicida*, *Francisella philomiragia*, *Francisella hispaniensis*, *Francisella guangzhouensis*, *Francisella haliotica*, et *Francisella noatunensis* (<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=262>) (Keim et al., 2007). Comme le montre la **Figure 1**, *Francisella tularensis* est divisée en trois sous-espèces, *Francisella tularensis* subsp. *tularensis*, *Francisella tularensis* subsp. *holarctica* et *Francisella tularensis* subsp. *mediasiatica*, sur la base de traits phénotypiques tels que la virulence.



**Figure 1** : Taxonomie de *Francisella* dont ses 3 principales espèces et les 3 sous-espèces de *Francisella tularensis*.

Tandis que les espèces *F. novicida* et *F. philomiragia* ont toujours été associées à une faible virulence, *F. tularensis* subsp. *tularensis* (type A) montre un fort taux de virulence se traduisant par un taux de mortalité pouvant atteindre 30 % lorsqu'il n'y a pas de traitement antibiotique disponible. Avec cette sous-espèce, l'exposition de 10 à 50 bactéries est

suffisante pour causer une infection intra-cutanée ou respiratoire (Foley and Nieto, 2010). Cette remarquable virulence est attribuée aux mécanismes de subversion que cet organisme a élaboré contre le système immunitaire des hôtes mammifères (Bosio, 2011). Par ailleurs, *F. tularensis* subsp. *holarctica* (type B) et *F. tularensis* subsp. *mediasiatica* ont une forte virulence mais aucun cas de mortalité humaine n'a été rapporté à ce jour (Keim et al., 2007).

### 1.3.1 Description de la bactérie

Tout d'abord décrite comme « *Bacterium tularense* » en 1911 par McCoy (McCoy, 1911), cette bactérie est transmise à l'homme par des morsures d'insectes, la manipulation de tissus infectés de rongeurs et de lapins, l'inhalation d'aérosols infectieux et l'exposition à de la nourriture ou à des boissons contaminées (Francis, 1925; Francis et al., 1922). C'est une bactérie coccobacille à Gram-négatif de dimension 0,1-1 x 0,1-3 µm, intracellulaire facultative, immobile, aérobique et non sporulée (Foley and Nieto, 2010). Les formes virulentes de *Francisella tularensis* sont entourées d'une capsule dont la perte n'affecte pas la viabilité mais s'accompagne d'une perte de virulence.

### 1.3.2 Mode de culture

*Francisella* est une bactérie difficile à cultiver : elle pousse lentement à 37°C et nécessite un milieu enrichi en cystéine ou en cystine. Les principaux milieux de culture et d'isolement sont les milieux de Francis, de McCoy et Chapin, glucose-cystéine-peptone, les géloses BCYE (Buffer Charcoal Yeast Extract) et chocolat enrichie en PolyVitex.

### 1.3.3 La maladie

La tularémie se manifeste par différents symptômes cliniques selon la dose, l'emplacement de la contamination et la virulence de la souche. Les formes principales sont les suivantes : ulcéro-ganglionnaire, ganglionnaire, oculo-ganglionnaire, typhoïde, pharyngo-ganglionnaire, pulmonaire (Dennis et al., 2001) et enfin une infection générale pouvant entraîner la mort (Bulut et al., 2013).

### 1.3.4 Répartition géographique naturelle

*Francisella* est majoritairement présente dans l'hémisphère nord. La prévalence géographique de certaines sous-espèces est reconnue : *F.t.* subsp. *mediasiatica* n'a été observée qu'en Asie centrale, l'observation de *F. t.* subsp. *tularensis* est restreinte à l'Amérique du nord alors que *F. t.* subsp. *holarctica* se trouve en Amérique du nord et en Eurasie (Gurycova, 1998; Keim et al., 2007).

### 1.3.5 Virulence

il a été estimé par des consultants de l'Organisation Mondiale de la Santé qu'un largage par avion de 50 kg de *F. tularensis* au-dessus d'une ville de 500 000 habitants entraînerait 30 000 décès et 125 000 invalides (Collectif, 1970).

La souche vaccinale de *F. tularensis*, *Francisella tularensis* subsp. *holarctica* LVS, est l'antigène du seul vaccin pour l'homme reconnu et disponible à l'heure actuelle (Conlan, 2011). Le vaccin n'est pas utilisé de manière courante car les deux modes d'administration

possibles ne montrent pas une efficacité totale. En effet, l'administration par voie cutanée ne protège pas suffisamment contre l'infection par voie respiratoire, et l'administration en aérosol montre trop d'effets secondaires pour être considérée comme une vaccination sûre. D'autres types de vaccins sont en cours de développement comme le décrit Conlan (Conlan, 2011) dans sa revue, mais il précise également que de nombreuses années de développement sont encore nécessaires avant de parvenir aux essais cliniques.

#### 1.4 Les méthodes de détection et d'identification existantes

Il existe à l'heure actuelle un large choix de méthodes permettant d'identifier et de détecter des microorganismes. La méthode idéale se doit d'être fiable, facile à mettre en œuvre, rapide, utilisable en haut débit et bien sûr à moindre coût (Ho and Reddy, 2010). Les différentes techniques à notre disposition pour identifier *Francisella tularensis* ainsi que leurs avantages et inconvénients inhérents sont détaillés ci-dessous. A l'heure actuelle, quatre méthodes sont principalement utilisées pour l'identification de l'agent causal de la tularémie : la culture, la PCR (Polymerase Chain Reaction, réaction de polymérase en chaîne), les tests sérologiques et la spectrométrie de masse (Hepburn and Simpson, 2008).

##### 1.4.1 Culture et tests phénotypiques

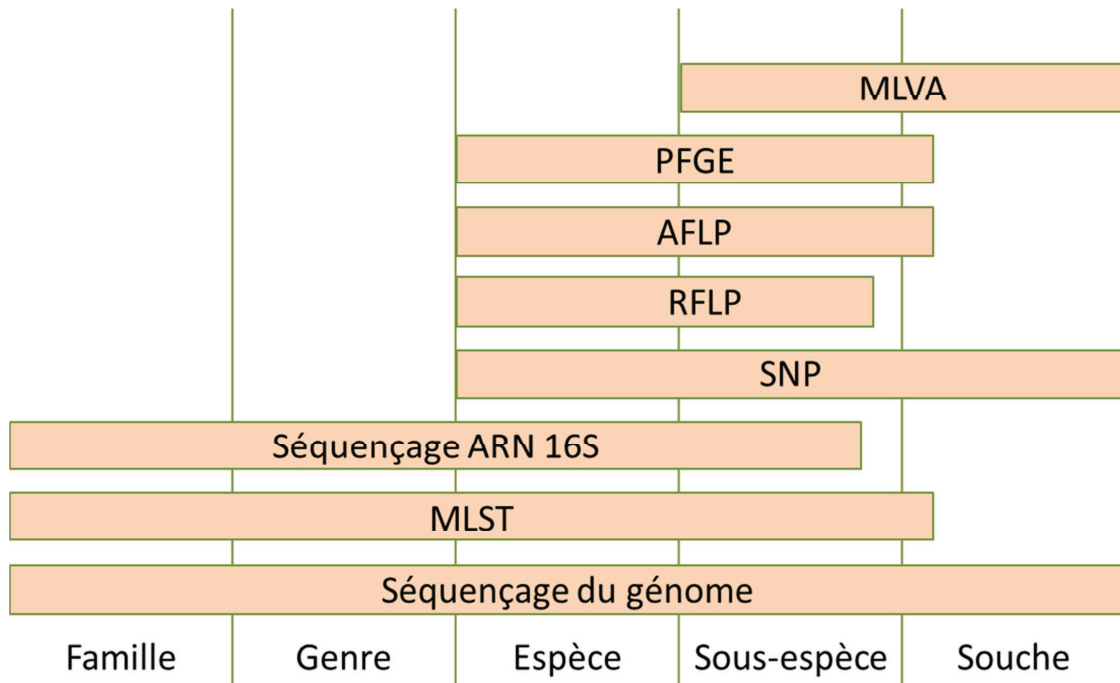
La culture microbienne est la méthode de référence. Elle permet d'identifier des bactéries vivantes et cultivables en analysant leurs caractères phénotypiques. Elle est complétée par les tests biochimiques réalisés grâce à la galerie API, dispositif regroupant un ensemble de tubes prêts à l'emploi. Pour l'identification de *Francisella* les critères biochimiques qui sont analysés en complément de la morphologie, la mobilité, la formation de l'endospore et la croissance en aérobiose ou en anaérobiose sont les suivants :

- la coloration de Gram (Gram-négative),
- la production de catalase (faible catalase),
- le test de l'oxydase (oxydase-négative pour *F. tularensis* et oxydase-positif pour *F. philomiragia*) (Johansson and Petersen, 2010),
- la formation d'acides ou de gaz à partir d'hydrates de carbone (catabolisme lent pour la production d'acide, pas de production de gaz),
- la production de sulfure d'hydrogène (positif),
- le test de détection chromogénique (positif),
- le test de l'uréase (négatif) (Johansson et al., 2012).

L'ensemble de ces critères d'identification est basé sur la culture bactérienne qui exige de bonnes connaissances en microbiologie. Elle est précise et fiable, mais elle est coûteuse en temps et en matériel et nécessite surtout que la bactérie soit vivante et cultivable.

### 1.4.2 Méthodes génomiques

De nombreuses méthodes d'identification génomique existent. Comme illustré sur la **Figure 2** elles permettent l'identification des *Francisella* à différents niveaux de résolution taxonomique. Ces méthodes se basent sur les différences de séquences de nucléotides qui existent dans les génomes au sein d'une même espèce et entre les espèces. L'analyse du génome complet des sous-espèces de *Francisella tularensis* montre que la diversité des séquences génomiques est très faible avec une moyenne d'identité des nucléotides de plus de 99,2 % (Johansson and Petersen, 2010).



**Figure 2** : Plage d'identification selon les méthodes génomiques. Inspiré de (Johansson et al., 2004).

- **Séquençage complet du génome** : cette méthode permet d'obtenir les meilleurs résultats pour une classification parfaite dans la phylogénie, que ce soit pour les organismes connus ou inconnus. Aujourd'hui cette méthode est toujours coûteuse en temps et en argent mais avec les avancées rapides de ces technologies nous pouvons imaginer du séquençage à haut débit et à moindre coût dans un futur proche.
- **Typage génomique multilocus (MLST)** : cette méthode repose sur le séquençage de fragments de 450 à 500 paires de bases (bp) de gènes de ménage. Pour le genre *Francisella* 7 gènes ont été choisis : *aroA*, *parC*, *pgm*, *tpiA*, *trpE*, *atpA* et *uup* (Svensson et al., 2005). Ces gènes permettent de distinguer plusieurs groupes de souches au sein des sous-espèces *tularensis* et *holarctica*, sans aller jusqu'à la distinction entre chaque souche.
- **Séquençage de l'ARN ribosomal (16S et 23S)** : Ces séquences d'ARNs sont très conservées au sein du genre et de l'espèce et permettent l'identification jusqu'au niveau sous-espèce de *Francisella* (Forsman et al., 1990).

- **Polymorphisme nucléotidique (SNP)** : pour cette méthode, ce sont les variations qui sont étudiées dans le génome entier. Selon le nombre de substitutions choisies les identifications peuvent avoir des résolutions atteignant le niveau souche. L'équipe de Vogler et al., en ciblant 23 SNP, a pu répartir 496 isolats de *F. tularensis* en 10 sous-populations de *F. t. holarctica* et 6 sous-populations de *F. t. tularensis* (Vogler et al., 2009a). Cette technique peut également mettre en évidence de nouvelles souches.
- **Polymorphisme de longueur des fragments amplifiés ou des fragments de restriction (AFLP ou RFLP)** : ces deux méthodes sont basées sur la digestion enzymatique du génome entier. L'AFLP permet, chez *Francisella*, de séparer à la fois les sous-espèces *tularensis* et *holarctica*, et chez *holarctica* de séparer les sous-populations (Garcia Del Blanco et al., 2002). Il est donc possible d'atteindre un niveau souche avec ces espèces, mais seulement dans des cas spécifiques.
- **Electrophorèse en champ pulsé (PFGE)** : le changement de direction du champ en fonction du temps permet de séparer les molécules d'ADN selon leur structure tridimensionnelle, séparant ainsi des fragments allant jusqu'à 10 Mb contre 50 kb avec une électrophorèse classique. Cette technique permet, après une digestion par des enzymes de macrorestriction de l'ADN total, de différencier les sous-espèces de *F. tularensis* et de les séparer en deux populations, mais il n'est pas possible d'aller jusqu'au niveau souche avec cette technique (Fey et al., 2007; Garcia Del Blanco et al., 2002; Staples et al., 2006).
- **Analyse multiple de séquences courtes répétées en tandem (MLVA)** : cette méthode se base sur l'analyse du nombre de répétitions de courtes séquences de nucléotides pour différents emplacements sur des chromosomes, appelés des loci. L'identification jusqu'au niveau souche est possible grâce à la grande variabilité du nombre de répétitions entre les souches. Chez *Francisella* ce nombre varie entre 2 et 20 sur 6 sites de séquences courtes répétées en tandem (STR) (Farlow et al., 2001; Vogler et al., 2009b)

#### 1.4.3 Méthodes immunologiques

Les deux méthodes immunologiques principales sont la micro-agglutination et les tests ELISA (Enzyme-Linked Immunosorbent Assay).

**L'agglutination ou la micro-agglutination** permettent de fixer les anticorps anti-*Francisella* présents dans le sérum à des antigènes de *Francisella*. C'est donc une méthode de détection indirecte de *Francisella*. Un phénomène d'agglutination va avoir lieu produisant un culot coloré si les antigènes sont marqués, ou une diminution de la densité optique (Massey and Mangiafi, 1974; Massey and Mangiafi, 1974; Sato et al., 1990). Mais en début d'infection où le titre d'antigène est faible, ces tests conduisent parfois à des faux-positifs (Hepburn and Simpson, 2008). Les tests **ELISA** sont possibles sur plusieurs types d'antigènes tels les lipopolysaccharides (LPS) ou d'autres antigènes extraits des bactéries. C'est une méthode de détection directe de *Francisella*. Le test ELISA peut être utilisé sur du sérum ou sur des échantillons environnementaux. De nombreuses études cherchent à miniaturiser ce type de

test afin d'en avoir une utilisation simple, rapide et systématique (Klemm et al., 2013). C'est dans le même esprit que Splettstoesser et al. développèrent un test bandelette immunochromatographique pour la détection d'anticorps dirigés contre *Francisella tularensis* dans le sérum (Splettstoesser et al., 2010).

#### 1.4.4 Méthodes de spectrométrie de masse

La spectrométrie de masse de type MALDI-TOF (Matrix Assisted Laser Desorption/Ionisation Time Of Flight) est devenue ces dernières années un nouvel outil bio-analytique de pointe pour l'identification et la détection de bactéries (Kliem and Sauer, 2012; Lay, 2001). Généralement les bactéries sont déposées entières sur des cibles métalliques et une solution de matrice y est ajoutée de manière à former des cristaux contenant les analytes. Des tirs laser permettent la désorption et l'ionisation des analytes. La mesure du temps de vol des protéines ou des peptides chargés à travers le tube de l'analyseur permet de calculer leur masse. L'identification se base ensuite sur l'existence de profils de masses de protéines ou de peptides, spécifiques à chaque microorganisme. L'analyse de ces profils se fait par l'intermédiaire de logiciels proposant des bases de données de spectres de référence (Demirev et al., 1999). Seibold et al. identifièrent correctement 45 souches de *Francisella* contre une base de spectres contenant 5 spectres de références de *Francisella* et 3 287 spectres d'autres organismes (Seibold et al., 2010). Cette méthode fournit des résultats similaires à ceux obtenus lors d'un séquençage de l'ARN 16S et 23S (Sauer and Kliem, 2010) et est désormais reconnue dans le diagnostic hospitalier. L'identification de biomarqueurs spécifiques par approche dite « bottom-up » (digestion enzymatique des protéines avant analyse des peptides permettant l'identification des protéines) est aussi possible même si elle est fastidieuse (Fagerquist et al., 2010; Sandrin et al., 2013). Pavkova et al. identifièrent ainsi 6 protéines exprimées uniquement chez la plus virulente des sous-espèces, *F. t. tularensis* (Pavkova et al., 2006).

## 2 Les apports de la protéomique et des nouveaux spectromètres de masse

La protéomique est la discipline qui étudie l'ensemble des protéines d'un organisme, d'un tissu, d'un organe, d'un fluide biologique, à un temps donné et sous des conditions données. La spectrométrie de masse, technologie permettant l'ionisation et la mesure du rapport masse sur charge ( $m/z$ ) de molécules, a été développée par le physicien J.J. Thomson au début du XX<sup>ème</sup> siècle. Elle fut d'abord utilisée pour mesurer les masses atomiques des éléments chimiques. Pendant la seconde guerre mondiale, les premières molécules chimiques sont analysées par spectrométrie de masse pour les besoins de l'industrie pétrolière (Yates, 2011). Ensuite, l'analyse des structures chimiques a pris son essor grâce à Klaus Biemann à la fin des années 1950. Dans les années 1960, deux avancées technologiques ont permis de développer le séquençage des peptides : l'ionisation chimique (Munson and Field, 1966) et le couplage de la chromatographie en phase gazeuse à la spectrométrie de masse (Gohlke, 1959; Ryhage, 1964). Dès la fin des années 1970



l'interprétation des spectres grâce à l'informatique s'automatise mais il n'est pas encore possible d'analyser des molécules plus grosses que des tri-peptides en raison des procédures de dérivation et de la difficulté à vaporiser des grands peptides (Yates, 2011). La spectrométrie de masse en tandem voit le jour grâce au développement du triple quadripôle et à Donald Hunt (Hunt et al., 1981). C'est l'arrivée des nouvelles sources d'ionisation qui va permettre en premier d'analyser de plus grosses molécules (quelques centaines de Daltons) avec le FAB (Fast Atom Bombardment) par Michael Barber et al. (Barber et al., 1981). Puis le MALDI (Matrix Assisted Laser Desorption/Ionization) développé par Michael Karas et Franz Hillenkamp (Karas and Hillenkamp, 1988), va permettre l'ionisation de larges molécules (protéines, glucides, ADN) et la mesure de leurs masses grâce à l'analyseur en temps de vol (TOF, Time-Of-Flight). Enfin l'ESI (Electrospray Ionization) développée par Fenn et al. (Fenn et al., 1989), va donner la possibilité de coupler la chromatographie en phase liquide aux spectromètres de masse, permettant ainsi l'analyse de fluides biologiques complexes par exemple. La protéomique prit son essor grâce à ces deux dernières sources d'ionisation. Rapidement il est apparu que les séparations de protéines par gel-2D (deux dimensions) suivies de digestions protéolytiques avant l'analyse par spectrométrie de masse, permettaient d'identifier facilement les protéines grâce à leur empreinte peptidique. L'innovation en matière d'analyseurs pour la spectrométrie de masse en tandem a augmenté la robustesse, la vitesse d'analyse, la résolution et la précision des masses.

## **2.1 Le « boom de la protéomique » en lien avec l'évolution de la spectrométrie de masse**

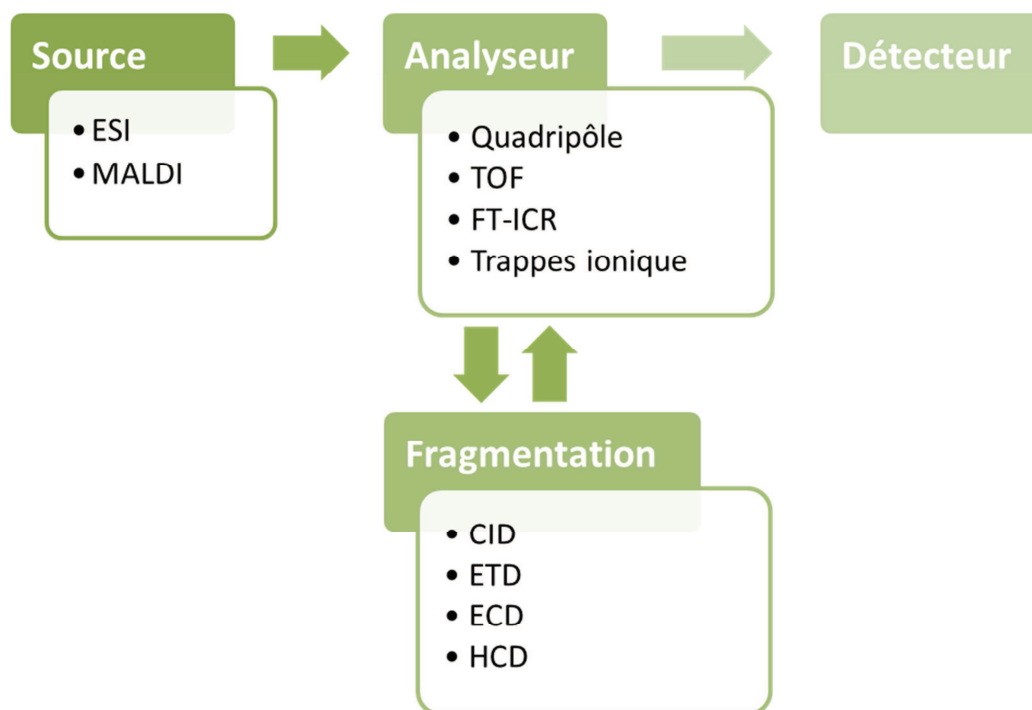
Les différentes techniques de spectrométrie de masse, décrites ci-dessus, ont permis de faire de grandes avancées dans l'étude des protéines. Pouvoir identifier mais également quantifier les protéines, à la fois résultats de l'expression des gènes et acteurs de l'activité cellulaire, est un enjeu majeur pour la compréhension des mécanismes cellulaires et pour le diagnostic (Aebersold and Mann, 2003).

La spectrométrie de masse (MS) permet l'étude très spécifique des protéines et notamment de leur structure primaire, des modifications post-traductionnelles (PTMs) et des interactions entre les protéines. Le développement des techniques de séparation et l'automatisation des machines permettent également d'étudier le protéome de manière globale. A l'heure actuelle, d'après GOLD (Genome Online Database, <http://www.genomesonline.org/>), ce sont plus de 35 000 génomes qui sont entièrement séquencés ou en cours de séquençage et disponibles sur le site du NCBI (<http://www.ncbi.nlm.nih.gov/genome/>). Ainsi la constitution de telles bases de données facilite l'identification des protéines et dans certains cas, la découverte et la correction d'erreurs d'annotation du génome.



## 2.2 La spectrométrie de masse

Ce paragraphe décrit les différentes parties qui composent un spectromètre de masse. Comme le montre la **Figure 3**, un spectromètre est composé de 3 éléments distincts. Tout d'abord la source qui génère les ions, ensuite un ou plusieurs analyseurs qui permettent de manipuler les ions sous vide et de faire la fragmentation le cas échéant, et enfin le détecteur qui mesure le signal produit par les ions.



**Figure 3** : Schéma général d'un spectromètre de masse. ESI : Ionisation par Electronébuliseur ; MALDI : Ionisation/Désorption Laser Assistée par Matrice. TOF : Temps de vol ; FT-ICR : résonance cyclotronique d'ions à transformée de Fourier. CID : Dissociation induite par Collision; ETD : Dissociation par Transfert d'Electron ; ECD : Dissociation par Capture d'Electron ; HCD : Dissociation induite à très haute énergie.

### 2.2.1 Les sources d'ionisation

En protéomique les deux sources principales d'ionisation sont des sources d'ionisation douces qui génèrent des ions relativement stables et dont la durée de vie est assez longue pour traverser l'analyseur et donc être détectés.

#### a. Ionisation par Electronébuliseur (ESI)

En source ESI, les molécules ne sont pas introduites simultanément. L'étape d'ionisation est souvent précédée d'une chromatographie en phase liquide sur colonne à séparation en phase inverse qui permet la séparation des peptides en fonction de leur hydrophobicité. Une fois séparés, les peptides passent par une aiguille à l'extrémité de laquelle, les gouttelettes sont vaporisées et ensuite sous l'action d'un fort champ électrique les peptides sont ionisés (Fenn et al., 1989). Il est également possible de faire de l'infusion de molécules, c'est-à-dire d'injecter une solution en continu sans séparation préalable.

### b. Désorption/Ionisation Laser Assistée par Matrice (MALDI)

En source MALDI, les analytes sont déposés sur une plaque métallique appelée cible et mélangés avec une matrice absorbant à une longueur d'onde donnée. En séchant, les analytes vont co-cristalliser avec la matrice. Les molécules de matrice vont absorber les photons UV transmis par le faisceau laser pulsé et ainsi s'exciter et s'ioniser. L'énergie absorbée provoque la dissociation de la matrice et son passage en phase gazeuse entraînant également les analytes. Les molécules de matrice ionisées transfèrent leur charge à l'analyte permettant ainsi leur détection (Hillenkamp et al., 1991).

#### i. Les différentes matrices

Différentes matrices chromophores peuvent être utilisées sur des cibles MALDI-TOF. Le choix de la matrice à utiliser se fait en fonction de la nature des molécules à analyser. Les solutions de matrices sont généralement préparées dans un mélange d'eau, de solvant organique (acétonitrile ou éthanol) et d'acide trifluoroacétique. Les matrices les plus utilisées en protéomique sont l'acide  $\alpha$ -cyano-4-hydroxycinnamique (HCCA) qui permet d'analyser des peptides et des protéines de masses inférieures à 20 kDa et l'acide sinapinique (SA) qui permet d'analyser des protéines entières de masses au-delà de 2 kDa. L'ensemble des matrices couramment utilisées selon le type d'échantillons est détaillé dans le **Tableau 1**.

**Tableau 1** : Matrices chromophores et leurs utilisations. SA : acide 3,5-diméthoxy-4-hydroxycinnamique (Acide Sinapinique) ; HCCA : acide  $\alpha$ -cyano-4-hydroxycinnamique ; HABA : acide 2-(4-hydroxyphenylazo)benzoïque ; MBT : 2-mercaptobenzothiazole ; DHB : acide 2,5-dihydroxybenzoïque ; THAP : 2,4,6-trihydroxyacetophenone, PA : Acide Picolinique ; HPA : acide 3-aminopicolinique ; DHAP : 2,6-dihydroxyacetophenone. Tiré de (Clark et al., 2013).

Matrices	Types d'échantillons
SA	Protéines
HCCA	Peptides, triacylglycérol
HABA	Peptides, protéines, glycoprotéines
MBT	Peptides, protéines, polymères synthétiques
DHB	Oligosaccharides
THAP	Oligonucléotides
PA, HPA	Oligonucléotides, ADN, biopolymères
DHAP	Glycopeptides, phosphopeptides

#### ii. Les modes de dépôts

La manière et l'ordre dans lesquels sont déposés les échantillons et la matrice peuvent avoir une influence sur l'homogénéité des dépôts et indirectement sur la qualité des spectres obtenus. Il existe trois modes de dépôt possibles. Le dépôt en « couche mince » est la technique la plus simple et la plus utilisée pour l'étude de microorganismes. Il s'agit de déposer en premier l'échantillon puis de déposer la matrice. Les analytes sont ainsi uniformément répartis mais le dépôt est fragile ce qui limite le nombre de tirs laser. La seconde méthode, dite de la « goutte séchée », consiste à mélanger extemporanément l'échantillon et la matrice avant de les déposer sur la cible. Les cristaux obtenus sont assez

grossiers mais cette méthode est très rapide à mettre en œuvre. La dernière méthode dite « en sandwich » consiste à déposer l'échantillon entre deux couches de matrice. Cette dernière méthode permet d'obtenir de petits cristaux réguliers ce qui va permettre d'augmenter la résolution et la qualité des spectres. Cette méthode est aussi plus délicate et coûteuse en temps à mettre en œuvre.

### 2.2.2 Les analyseurs

Une fois à l'intérieur du spectromètre de masse les peptides ionisés sont dirigés et manipulés par des champs électriques sous vide. Pour déterminer les rapports masse sur charge ( $m/z$ ), il existe quatre types d'analyseurs : les quadripôles, les temps de vol (TOF), les analyseurs à résonance cyclotronique d'ions à transformée de Fourier (FT-ICR) et les trappes ioniques (tridimensionnelle, linéaire et Orbitrap). Ces analyseurs peuvent aussi être couplés entre eux pour améliorer leurs performances. Ces instruments génèrent des spectres de masses donnant l'intensité des ions en fonction de leur rapport  $m/z$ .

- **Le quadripôle** est constitué de quatre électrodes cylindriques et parallèles auxquelles est appliqué un potentiel sinusoïdal permettant de focaliser les ions. Le quadripôle agit comme un filtre de masse en sélectionnant les ions à un ratio masse sur charge donné.
- **Les analyseurs à temps de vol (TOF)** mesurent le temps mis par les ions pour parcourir une distance donnée (la longueur du tube de vol). Ce sont des analyseurs à haute résolution. A partir du temps de vol le rapport masse sur charge est directement calculable d'après la formule de l'énergie cinétique (**Équation 1**).

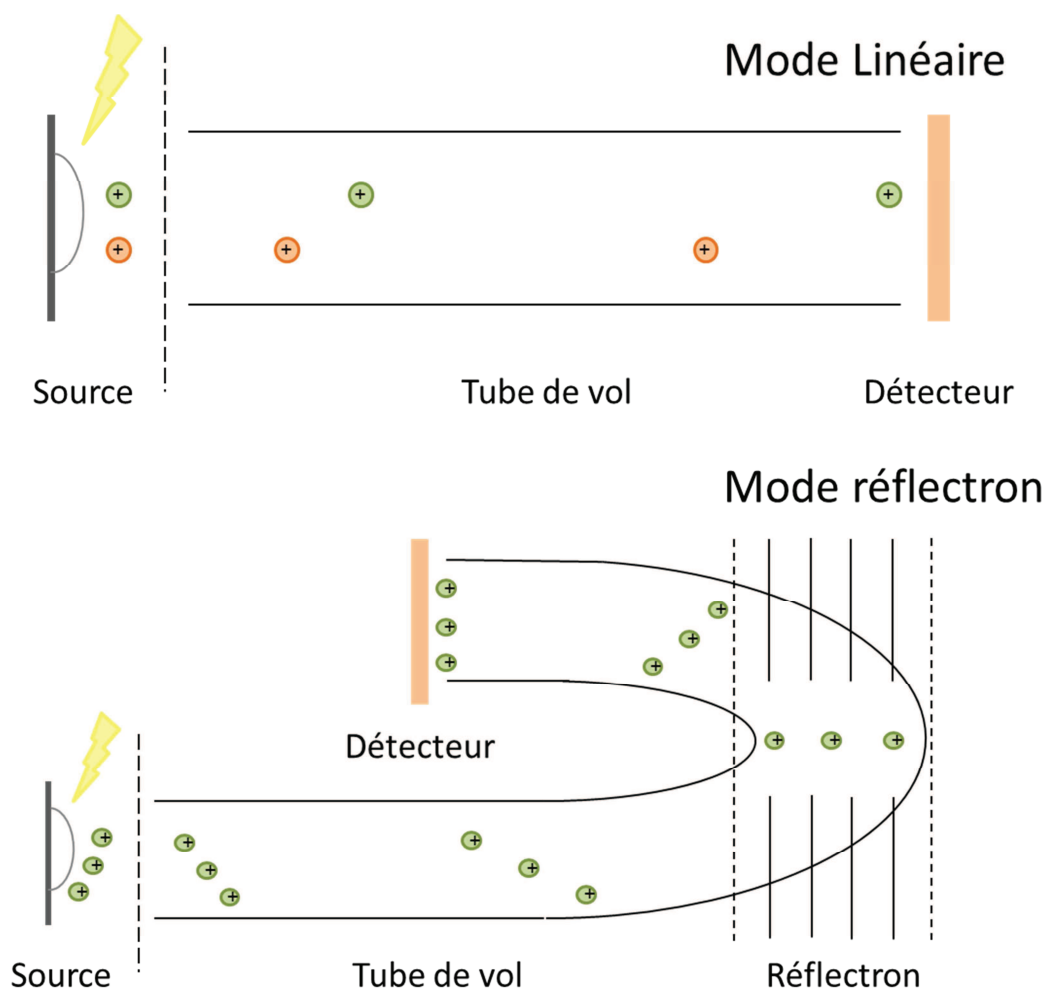
**Équation 1** : Calcul du rapport masse sur charge

$$Ec = \frac{1}{2}mv^2 = \frac{1}{2}m\left(\frac{l^2}{t^2}\right) = zVo$$

$$\frac{m}{z} = 2Vo\frac{t^2}{l^2}$$

- |                                 |                                |
|---------------------------------|--------------------------------|
| - $Ec$ : énergie cinétique      | - $t$ : temps de vol           |
| - $m$ : masse de la protéine    | - $z$ : charge de la protéine  |
| - $v$ : vitesse                 | - $Vo$ : tension accélératrice |
| - $l$ : longueur du tube de vol |                                |
- 
- Les analyseurs à temps de vol sont constitués d'une zone d'accélération puis du tube de vol. Deux modes d'analyses sont possibles tels qu'indiqués sur la **Figure 4**. Le mode linéaire est utilisé pour les plus grosses molécules comme les protéines entières. Les ions sont détectés au bout du tube de vol. Le second mode, le mode

réflectron, est préféré pour des petites molécules de type peptides. Les ions sont freinés à l'extrémité du tube de vol par un miroir électrostatique, symbolisé par des lignes verticales sur la **Figure 4**, en imposant un champ électrique inverse par rapport à celui de l'entrée de l'analyseur. Les ions en ressortent avec une vitesse longitudinale de sens opposé à leur vitesse initiale. Les ions parcourent donc deux fois la longueur du tube de vol avant d'être détectés, assurant ainsi une meilleure résolution sans augmenter la taille de l'analyseur.



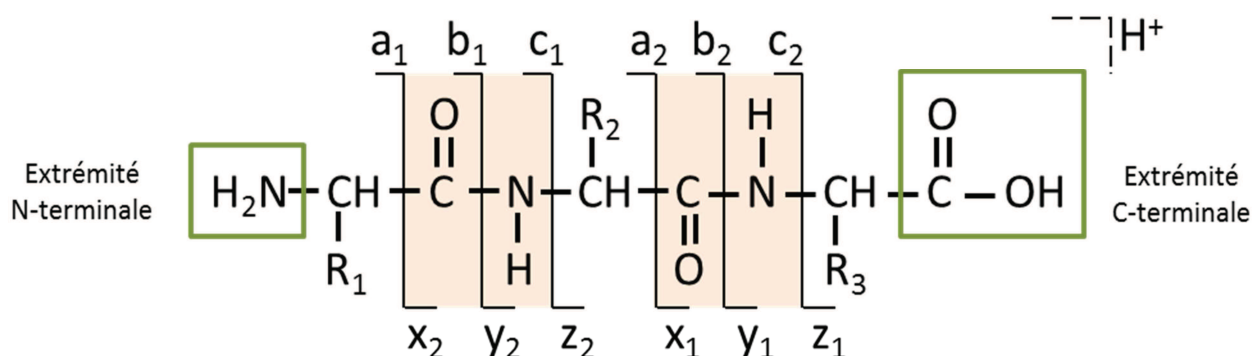
**Figure 4** : Mode linéaire et mode réflectron du tube de vol

- **Les analyseurs à résonance cyclotronique d'ions à transformée de Fourier (FT-ICR).** Dans ces analyseurs les ions sont piégés dans la cellule ICR ou piège de Penning. Ils sont soumis à un champ magnétique et subissent de ce fait une force de Lorentz qui leur donne un mouvement de rotation orthogonal au champ magnétique. La fréquence de rotation dépend du champ magnétique et du rapport  $m/z$ . La transformée de Fourier permet de transformer le spectre de fréquences en spectre de masses.

- **Les trappes ioniques** stockent les ions, permettant ainsi de mesurer leur masse et aussi de les isoler et de les fragmenter. Il existe plusieurs géométries de trappes : les trappes linéaires et tridimensionnelles qui sont de basse résolution, et l'Orbitrap qui est de haute résolution. Elles sont constituées de deux électrodes chapeau et d'une électrode annulaire. L'application d'une radiofréquence sur l'électrode annulaire produit un champ quadripolaire au sein duquel chaque ion acquiert un mouvement oscillant dont l'amplitude et la fréquence dépendent du rapport  $m/z$  de l'ion et de la valeur de tension appliquée. Sur l'électrode annulaire, l'augmentation de cette tension va provoquer l'éjection des ions selon leur rapport  $m/z$  pour être détectés.

### 2.2.3 Le principe de fragmentation

L'objectif de la fragmentation est de générer à partir de peptides de plus petits fragments pour permettre une analyse précise de la séquence des peptides par spectrométrie de masse. Une protéine est une séquence d'acides aminés reliés entre eux par des liaisons peptidiques dont l'extrémité N-terminale porte une amine primaire et l'extrémité C-terminale, un acide carboxylique. Ce sont ces différences qui définissent le sens de lecture de la séquence d'acides aminés (**Figure 5**).

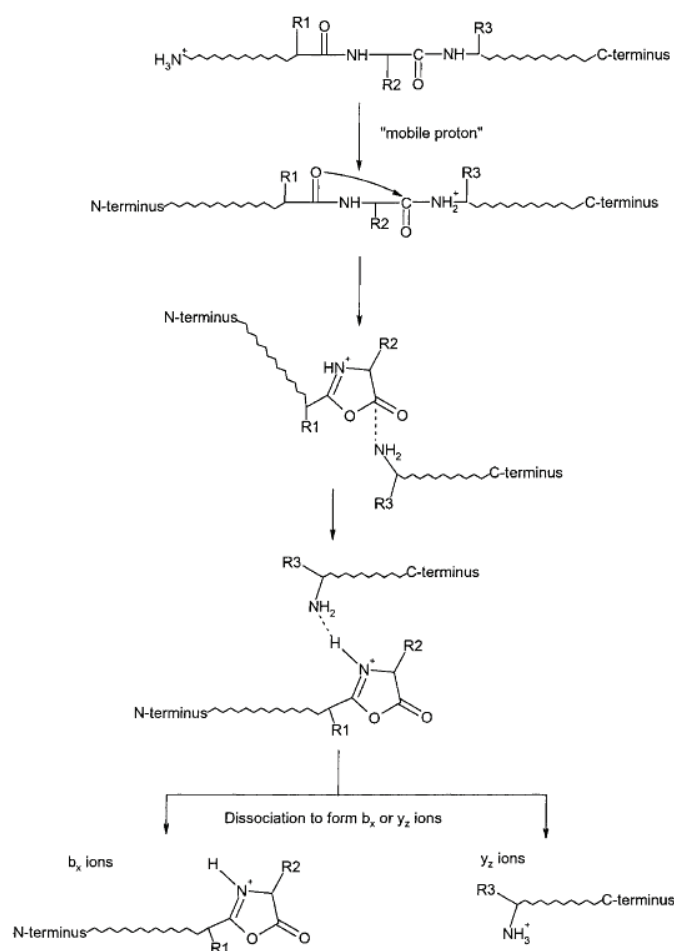


**Figure 5** : Nomenclature peptidique. Entourées de vert sont représentées les extrémités du peptide, les liaisons peptidiques sont en orange et le nom des différents ions fragments sont indiqués au-dessus et en-dessous de la séquence.

Pour générer des fragments peptidiques, on procède à la protéolyse des protéines. La protéolyse consiste à rompre les liaisons peptidiques (**Figure 5**) de manière spécifique en utilisant des protéases. L'enzyme la plus couramment utilisée est la trypsine, qui coupe en C-terminal des lysines (K) et des arginines (R), deux résidus basiques, facilitant ainsi la fragmentation en spectrométrie de masse en tandem (Olsen and Mann, 2004). On note également l'utilisation moins fréquente mais possible de la chymotrypsine (rupture en C-terminal des tyrosines (Y), tryptophanes (W), phénylalanines (F) ou leucines (L)), de l'endoprotéase Glu-C (rupture en C-terminal des acides glutamiques (E)) et de l'endoprotéase Lys-C (rupture en C-terminal des lysines (K)).

La détection des peptides par les analyseurs des spectromètres de masses permet la génération de spectres donnant l'intensité des peptides en fonction de leur rapport  $m/z$ .

Une fois ce rapport obtenu pour un peptide, le spectromètre de masse peut déterminer sa structure primaire, c'est-à-dire la séquence d'acides aminés du peptide. La spectrométrie de masse en tandem (MS/MS) consiste à isoler l'ion peptidique à un rapport  $m/z$  donné. Cet ion précurseur est alors fragmenté par collision avec un gaz inerte (argon ou hélium par exemple) pour donner des ions « fils » (Steen and Mann, 2004). La fragmentation va rompre les liaisons peptidiques comme le détaille la **Figure 6** selon la théorie du proton mobile. La charge pouvant se placer sur l'un ou l'autre des fragments, un fragment neutre et un fragment chargé sont obtenus: l'ion « fils ». Si la charge est sur le fragment N-terminal, c'est un ion b, si elle est sur le fragment C-terminal, ce sera un ion y (Roepstorff and Fohlman, 1984), (**Figure 5**).

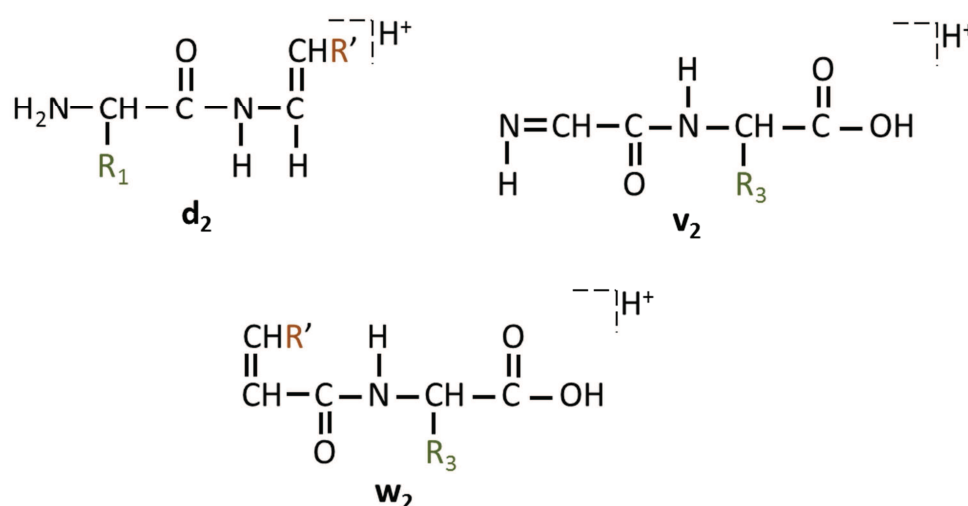


**Figure 6 :** Théorie du proton mobile. Mécanisme de fragmentation classique décrit par Paizs et Suhai (Paizs and Suhai, 2005). Le proton additionnel est mobilisé par l'atome d'azote de la liaison amide concernée par la fragmentation. L'attaque nucléophile de l'atome d'oxygène du groupement carbonyle voisin sur le carbonyle de la liaison amide concernée forme un dérivé cyclique oxazolone protoné avec une rupture de la liaison amide en son milieu. Soit la charge reste sur le dérivé d'oxazolone et c'est l'ion fragment b qui est détecté. Soit une modification conformationnelle peut provoquer un transfert de proton d'un fragment à l'autre et c'est alors l'ion fragment y qui est détecté (Rusconi, 2011).

### 2.2.4 Les méthodes de fragmentation

Pour faire de la spectrométrie de masse en tandem ( $MS^2$  ou  $MS/MS$ ), il faut fragmenter les peptides de manière à obtenir des ions fragments qui seront analysés par la suite. La méthode de fragmentation est choisie en fonction du type d'appareillage et de l'objet d'étude.

- **CID** : Collision Induced Dissociation. La fragmentation a lieu sous l'effet de la collision des ions avec un gaz injecté dans la chambre de collision selon l'énergie utilisée. La fragmentation se produit préférentiellement sur la liaison amine de la chaîne principale du peptide. En basse énergie (eV), les fragments a, b et y, indiqués sur la **Figure 5** sont observés (présence majoritaire des ions b et y). En haute énergie (keV), il est également possible d'observer les fragments d, v et w, dont les structures sont présentées sur la **Figure 7** (Johnson et al., 1988). Ces ions sont issus de la rupture de la chaîne peptidique et de la chaîne latérale de l'acide. Les ions d et w permettent notamment de différencier les acides aminés isobares comme la leucine et l'isoleucine.



**Figure 7** : Structure des ions fragments d, v et w.

- **ETD** : Electron-Transfer Dissociation. Ce mode de fragmentation se produit dans une trappe ionique et permet l'analyse des modifications post-traductionnelles ou des peptides très basiques. La fragmentation se produit par le transfert d'un électron provenant d'un anion radicalaire au peptide protoné. Les ions fragments c et z sont majoritairement observés à la différence du CID où ce sont les ions fragments b et y (Mikesh et al., 2006).
- **ECD** : Electron Capture Dissociation. Ce mode de fragmentation s'applique en FT-ICR-MS. Un électron à basse énergie réagit avec le peptide protoné dans le champ magnétique du FT-ICR. Ceci produit un peptide contenant une charge supplémentaire. Avec cette méthode de fragmentation les PTMs labiles telles que les

phosphorylations, N-glycosylations, O-glycosylations, sulfonations, etc. restent intacts (Zubarev et al., 1998).

- **HCD** : Higher-energy Collisional Dissociation. Ce mode de fragmentation est spécifique aux instruments de type Orbitrap. La fragmentation a lieu dans une cellule de collision externe. Ce type de fragmentation favorise l'observation des ions fragments de faibles masses (ions immoniums) et facilite le séquençage *de novo* (Olsen et al., 2007).

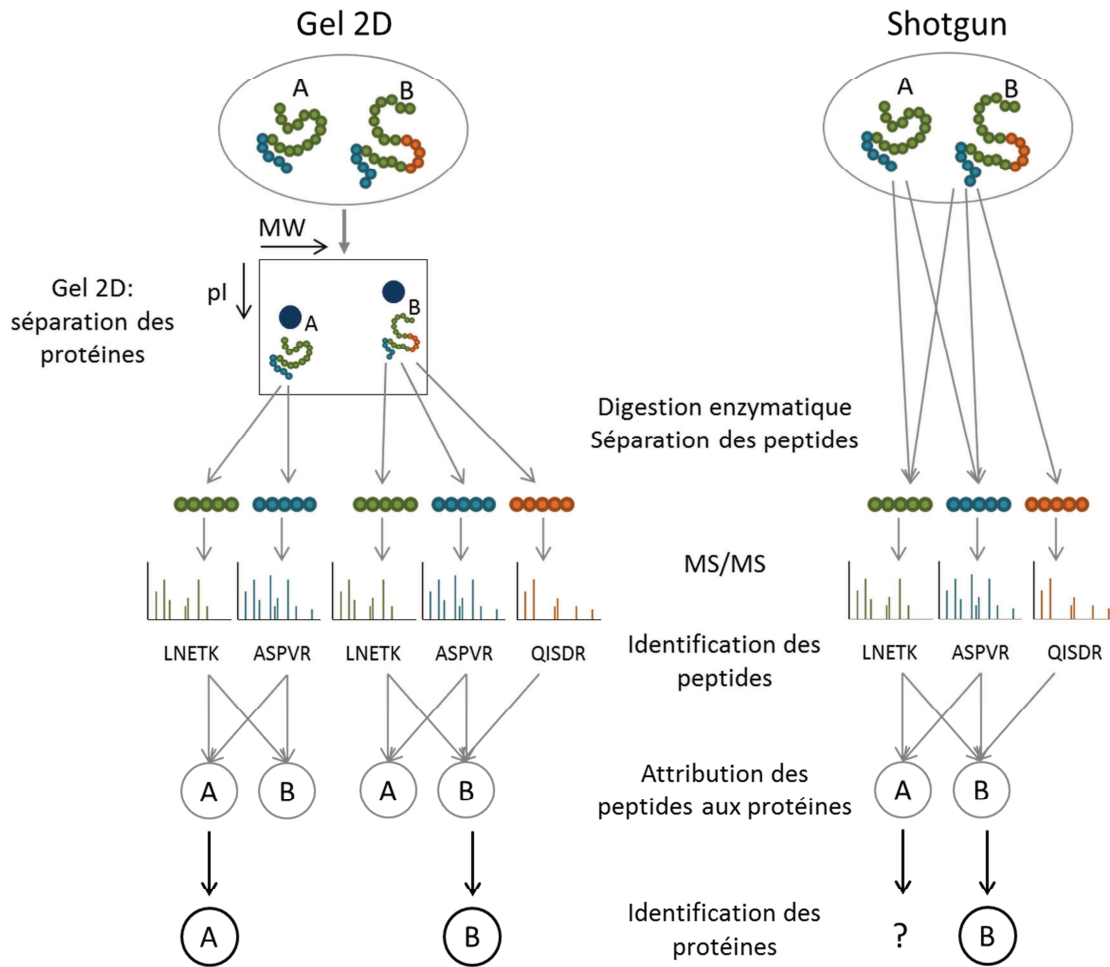
## 2.3 Shotgun et quantification, les nouvelles approches

Les quantités de données obtenues par spectrométrie de masse vont croissantes avec les avancées technologiques. L'objectif de la protéomique étant de donner un sens biologique aux données acquises nous allons ici développer les différentes approches en vigueur pour traiter ces informations.

### 2.3.1 La technique shotgun

La stratégie shotgun est basée sur la digestion enzymatique de protéines en peptides puis sur l'analyse de ceux-ci par spectrométrie de masse et enfin sur leur identification par recherche dans des bases de données. Contrairement à la séparation sur gel-2D, où les protéines sont digérées après avoir été séparées et sont donc analysées une par une, la méthode shotgun permet d'analyser un échantillon complexe où toutes les protéines sont digérées en même temps. Cette fois-ci, ce sont les peptides qui sont séparés avant l'analyse par spectrométrie de masse. Cette technique favorise l'obtention d'un jeu de données plus large qu'avec une séparation préalable sur gel-2D. En revanche comme le montre la **Figure 8**, l'interprétation des données peut se révéler plus complexe. En effet, notamment chez les eucaryotes, les protéines peuvent partager des peptides, ce qui peut conduire à des ambiguïtés lors de l'identification des protéines (Nesvizhskii and Aebersold, 2005).





**Figure 8 :** Identification de protéines par MS/MS. En gel 2D, les protéines sont préalablement séparées en fonction de leur masse (MW) et de leur point isoélectrique (pI). Après digestion enzymatique les protéines peuvent être identifiées sans ambiguïté. Avec la technique shotgun, toutes les protéines sont digérées simultanément et les peptides sont généralement séparés sur une nanoLC avant l'analyse par spectrométrie de masse. Si des peptides appartiennent à plusieurs protéines, comme ci-dessus avec un partage des peptides entre les protéines A et B, une ambiguïté peut se produire lors de l'identification des protéines s'il n'y a pas de présence de peptide spécifique à chaque protéine. Ici la protéine A n'a pas de peptide spécifique, il est donc impossible de conclure sur sa présence ou son absence dans l'échantillon. Inspiré de (Nesvizhskii and Aebersold, 2005).

Un grand nombre d'améliorations sont apportées sur les appareils de spectrométrie de masse avec comme objectif de pouvoir améliorer la technique shotgun en séquençant tout ce qui est présent dans l'échantillon :

- Amélioration de la séparation par chromatographie en phase inverse avec des colonnes plus longues, des particules de support de plus petite taille et des gradients plus longs, des nano-débits et des systèmes sous haute pression.
- Amélioration de la sensibilité et de la vitesse de traitement des spectromètres de masse, avec une meilleure injection des ions, des optiques de transfert plus performantes et une accélération de la vitesse d'acquisition des spectres.

- Amélioration de la résolution et de la précision des masses permet de faciliter les identifications par la suite, de même que l'utilisation de modes de fragmentation complémentaires (Altelaar et al., 2013).

Comme illustration de ces avancées, le protéome complet d'une levure a pu être obtenu en 2012 en une seule analyse par spectrométrie de masse d'environ 240 minutes par rapport à plusieurs jours d'acquisitions nécessaires en 2008 (de Godoy et al., 2008; Nagaraj et al., 2012).

### 2.3.2 Méthodes de quantification

Deux méthodes sont appliquées pour la quantification des protéines par spectrométrie de masse.

La quantification dite « label-based » repose sur le marquage chimique ou métabolique des protéines, à différents niveaux du traitement des échantillons. En effet, il est possible de marquer les protéines lors de la culture, après l'extraction des protéines avant la digestion ou après la digestion. Cette technique peut entraîner des erreurs de quantifications liées aux différences de traitement qu'il peut y avoir entre les échantillons marqués et non marqués. Plusieurs méthodes existent :

- le SILAC (Stable Isotope Labeled Aminoacids in Cell culture) consiste à marquer les protéines au moment de la culture en ajoutant dans le milieu des isotopes stables lourds (Ong et al., 2002).
- l'ICAT (Isotope Coded Affinity Tags) consiste à marquer les protéines sur les fonctions thiol des cystéines par des isotopes lourds ou légers. Après digestion trypsique les peptides sont séparés par une chromatographie d'affinité. Les peptides lourds et légers élués ensemble sont distingués par spectrométrie de masse (Gygi et al., 1999).
- l'iTRAQ (isobaric Tags for Relative and Absolute Quantification) consiste à marquer les amines primaires des peptides tryptiques avec des étiquettes isobariques pour une quantification relative. Les échantillons sont mélangés avant l'analyse par spectrométrie de masse et la quantification se fait en comparant les intensités des ions fragments de chaque étiquette (Ross et al., 2004).

Les stratégies dites « label-free » sont des stratégies sans marquage. Une quantification relative est possible à travers le nombre de spectres (spectral count) et/ou l'intensité des pics (extraction des chromatogrammes d'ions, XIC). Si la méthode du spectral count a l'avantage d'être simple à mettre en œuvre puisqu'il s'agit de compter le nombre de spectres attribué à une protéine, elle présente de nombreux inconvénients. En effet, cette méthode ne tient pas compte du fait que certains peptides peuvent être attribués à plusieurs protéines, ni que certaines protéines peuvent être identifiées avec peu de peptides (courte séquence polypeptidique, propriété physico-chimique des peptides altérant l'observabilité des peptides, erreur d'attribution, complexité du mélange, faible concentration de la protéine). La méthode du XIC repose sur l'extraction des

chromatogrammes des ions pour calculer l'intensité de signal proportionnelle à l'abondance du peptide parent. Une fois le temps de rétention du peptide connu, il est associé à la valeur  $m/z$  de l'ion précurseur et le chromatogramme de l'ion peut être extrait. La comparaison des aires des pics des peptides permet la quantification relative. Pour pouvoir extraire les chromatogrammes, la chromatographie doit être stable afin qu'il n'y ait pas de variation dans les temps de rétention (Mueller et al., 2008).

Avec le développement des biomarqueurs, l'ajout de peptides spécifiques (Absolute, Quantification, AQUA) ou de peptides concaténés (concatemer of standard peptides for absolute quantification, QconCAT) marqués par des isotopes lourds avant l'analyse par spectrométrie de masse permettent de quantifier de manière absolue. Les désavantages de ces deux méthodes sont les suivants : i) comme ce sont des peptides, l'efficacité de la protéolyse n'est pas prise en compte ; ii) dans le cas de matrices complexes, ils ne peuvent pas être intégrés à l'étape de pré-fractionnement le cas échéant ; iii) l'unique peptide marqué choisi n'est pas forcément représentatif de l'ensemble des peptides de la protéine à quantifier. La méthode PSAQ (Protein Standard Absolute Quantification) développée par l'équipe de Brun évite ces écueils en travaillant sur l'ajout dans les échantillons à quantifier des protéines entières marquées par des isotopes lourds (Brun et al., 2007; Picard et al., 2012)

La SRM (Selected Reaction Monitoring) est une technique montante pour la quantification, notamment pour étudier les variations d'expression de protéines. Cette technique exige que la protéine recherchée possède au moins un peptide protéospécifique et nécessite de connaître la transition correspondant au  $m/z$  du peptide et au  $m/z$  de l'ion fragment dont l'intensité va être mesurée pour la quantification. L'étude de plusieurs transitions (MRM, Multiple Reaction Monitoring) pour une même protéine augmente la spécificité analytique. Cette technique permet aussi de faire de la quantification absolue en ajoutant des isotopes lourds des peptides ciblés qui jouent un rôle de standard interne. Cette méthode voit ses limitations dans les échantillons complexes et dans les erreurs de mesures inhérentes aux instruments de spectrométrie de masse (Altelaar et al., 2013).

De nouvelles méthodes moins contraignantes que la SRM ont vues le jour récemment. Le SWATH-MS consiste à acquérir en haute résolution tous les ions fragments de rapport  $m/z$  compris entre 400 et 1 200 en effectuant un balayage au cours du temps. L'analyse se base ensuite sur la sélection des ions fragments pertinents contenus dans les bases de données. Les ions fragments co-élus peuvent être associés de manière à identifier les peptides à partir de leurs masses attendues. Pour cela, et c'est la limitation actuelle de cette méthode, des bases de données spécifiques doivent être créées pour chaque type d'échantillons. Le grand avantage de la SWATH-MS est l'acquisition de la totalité des données et non pas le suivi d'une transition ciblée, laissant ainsi la possibilité de retraiter les informations (Gillet et al., 2012).

### 3 Interprétation des spectres MS et MS/MS

Les spectromètres de masse génèrent une importante quantité de données qu'il est difficile de traiter de manière manuelle. Un grand nombre de logiciels permettent de traiter automatiquement ces données selon les besoins de l'utilisateur. Les algorithmes génèrent différents types d'indicateurs statistiques afin de permettre à l'utilisateur d'interpréter correctement les données obtenues.

#### 3.1 Attribution de séquences peptidiques aux spectres MS/MS

L'identification des protéines après digestion protéolytique (bottom-up) se fait de manière indirecte en identifiant les peptides composant ces protéines. En effet, chaque spectre MS/MS a été généré à partir d'un peptide. Il existe différentes approches d'attribution de séquences peptidiques : celles basées sur une base de données protéiques, le séquençage *de novo*, ou la recherche d'étiquettes.

##### 3.1.1 Attribution à partir d'une base de données – Le logiciel Mascot

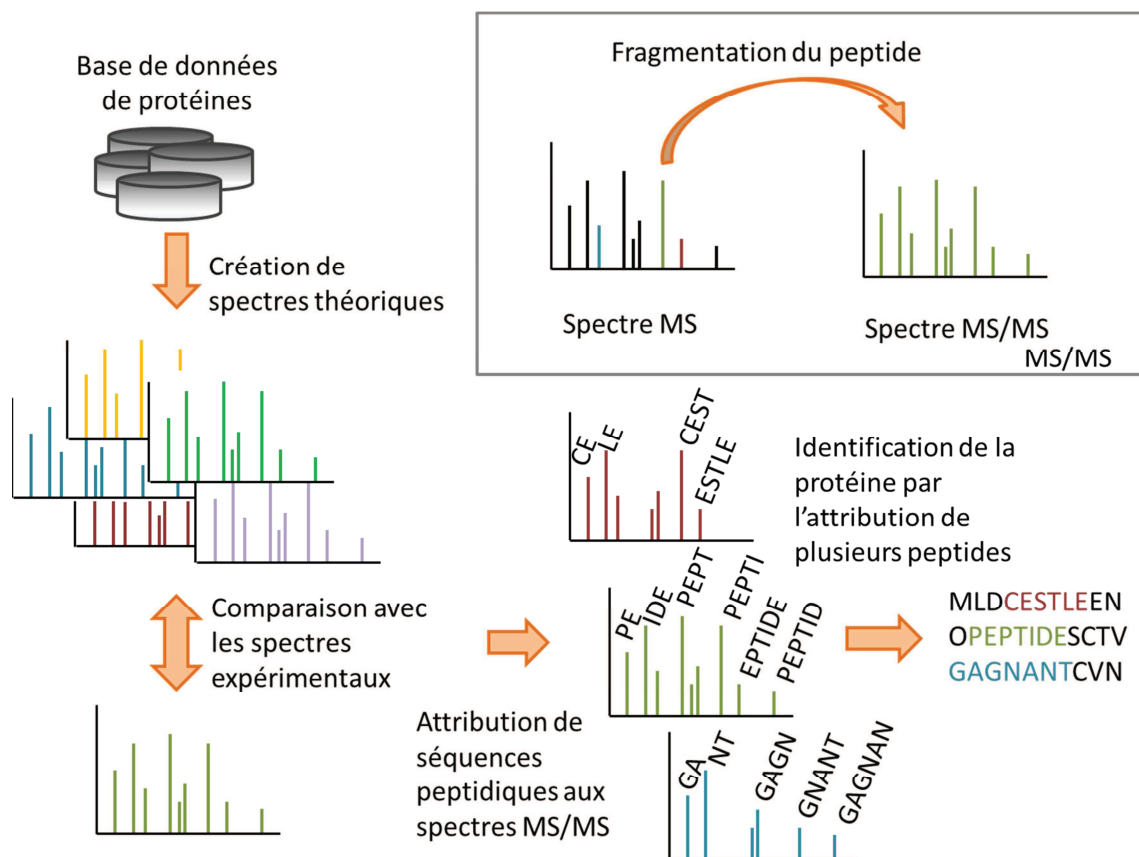
L'ensemble des séquences protéiques connues est regroupé dans des bases de données. Ces bases de données sont très variées en termes de spécificités d'organismes et de techniques d'annotations. Par exemple, la base de données Swiss-Prot ne présente que des séquences qui ont été validées manuellement à l'inverse de bases de données comme GenBank où les séquences peuvent être présentes à plusieurs reprises et sous différents noms. La base NCBI nr est une base de données qui contient des séquences uniques de protéines provenant des bases Swiss-Prot, GenBank, Protein Information Ressource (PIR), Protein Research Foundation (PRF) et Protein Data Bank (PDB).

Le principe d'attribution de séquences peptidiques aux spectres MS/MS est le suivant. Les spectres MS/MS obtenus expérimentalement sont comparés aux spectres théoriques issus de la base de données utilisée par un logiciel d'attribution, comme le détaille la **Figure 9**. Pour chaque spectre MS/MS, ce logiciel a pour but d'identifier la séquence du peptide qui a généré ce spectre parmi les séquences peptidiques issues de la digestion théorique des protéines de la base de données choisie. Le logiciel peut attribuer plusieurs séquences peptidiques à un spectre MS/MS étant donné l'incertitude qui persiste sur la nature de la séquence peptidique du spectre. Une attribution est appelée Peptide Spectrum Match (PSM). Les logiciels les plus utilisés actuellement sont Mascot (Matrix Science), SEQUEST (Thermo Scientific) et X!Tandem (The Global Proteome Machine) (MacCoss et al., 2002; Perkins et al., 1999). Chaque logiciel possède son propre algorithme. Par conséquent, chaque logiciel délivre un score qui lui est propre à chaque PSM.

Mascot, Le logiciel utilisé au laboratoire, calcule un score à chaque PSM qui est basé sur la probabilité  $P$  que l'attribution d'une séquence peptidique de la base de données à un spectre soit un hasard. Les modalités de calcul de cette probabilité  $P$  n'est pas divulguée par les auteurs de Mascot. Le score est calculé comme  $-10 \log(P)$ . Une très forte probabilité entraîne un faible score, il est donc plus probable qu'il s'agisse d'une erreur, alors qu'une faible probabilité donne un score important, l'attribution est donc de confiance. Il s'agit

d'identifier un score seuil à partir duquel les identifications sont considérées comme correctes. Ce score seuil est un compromis entre la sensibilité (taux de vrais positifs) et la spécificité (taux de faux positifs) désirées. Le False Discovery Rate (FDR) permet d'évaluer la quantité de peptides identifiés de manière incorrecte. Généralement une stratégie de recherche « target-decoy » est proposée pour évaluer le FDR (Elias and Gygi, 2010). Il s'agit de confronter le PSM à la base de données de protéines et en parallèle de le confronter à une base de donnée leurre qui est un mélange aléatoire de la base de donnée originale. Ceci permet de conserver la même nature et la même quantité d'information dans la base de données leurre que dans la base de données originale. Le nombre d'identifications dans cette base de données aléatoire permet d'évaluer le FDR.

Enfin, les protéines de l'échantillon sont identifiées à partir des séquences peptidiques identifiées avec confiance. Cette étape d'inférence des protéines est délicate (Nesvizhskii et al., 2005), de nombreuses règles ont été fixées afin de standardiser les pratiques de la communauté scientifiques (Carr et al., 2004). Le logiciel Mascot se base le principe de parcimonie. Il fournit le plus petit ensemble de protéines qui explique les séquences peptidiques identifiées. En effet, si pour quatre séquences, deux sont attribuées à la protéine A et deux sont attribuées à la protéine B, mais que les quatre sont trouvées dans la protéine C alors Mascot retournera comme résultat que c'est la protéine C qui est identifiée. Les protéines A et B feront parties du sous-ensemble des protéines identifiées avec ces séquences.



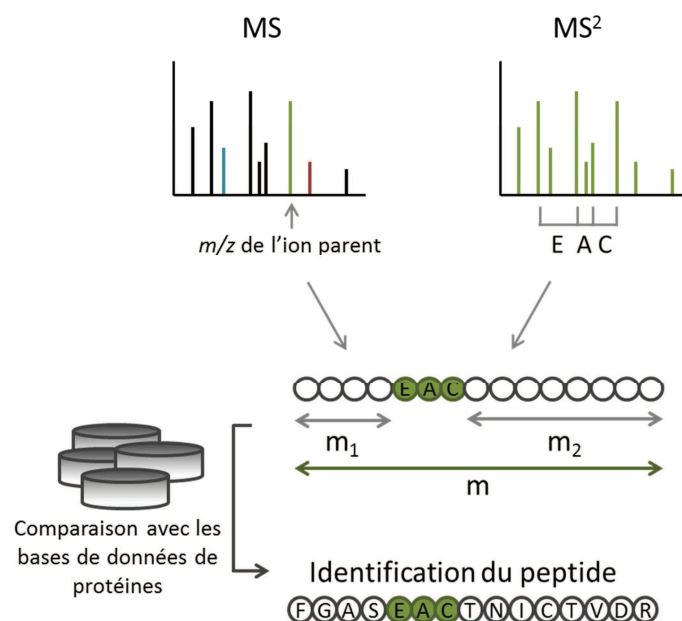
**Figure 9** : Fonctionnement général d'un algorithme d'identification de protéines

### 3.1.2 Séquençage de novo

Lorsque la masse de l'ion précurseur n'est pas assignée à une protéine particulière ou que la séquence du polypeptide est inconnue, il est possible, à partir du spectre obtenu en MS<sup>2</sup>, de reconstituer la séquence d'acides aminés du peptide. C'est le séquençage *de novo*. Ceci peut être fait de manière automatique, à l'aide de logiciels tels que PEAKS® (BSi), PepNovo (Frank, 2009a, b) ou Lutfisk (Taylor and Johnson, 1997), ou manuellement. Ce travail est lent et difficile et doit s'appuyer sur les meilleures résolutions et précisions des spectromètres de masse les plus récents. En effet, même s'il existe des règles de fragmentation que les logiciels modélisent pour faciliter l'analyse des spectres, la nature variée des acides aminés (acide, basiques, hydrophobes...) et les interactions qu'il peut y avoir entre eux, ainsi que la présence de modifications post-traductionnelles, imposent une vérification manuelle des annotations pour confirmer ou non les séquences établies automatiquement.

### 3.1.3 Recherche d'étiquette

Une dernière méthode existe, se plaçant entre le séquençage *de novo* et l'attribution de séquences peptidiques à des spectres MS/MS provenant de bases de données. Il s'agit de la recherche de signature d'étiquette. Par séquençage *de novo* une courte séquence d'acides aminés (3 à 5 AAs) est identifiée dans une zone interprétable du spectre comme montré sur la **Figure 10**. Ensuite cette séquence est encadrée par les masses aux deux extrémités et c'est cet ensemble qui est confronté aux bases de données pour identifier les peptides. Plusieurs logiciels procèdent ainsi, comme SPIDER (BSi) ou GutenTag (Tabb et al., 2003) par exemple (Nesvizhskii et al., 2007).



**Figure 10** : Identification par recherche de signature d'étiquette. La masse du peptide est déterminée à partir du spectre MS du peptide. A partir du spectre de fragmentation (MS<sup>2</sup>), une séquence d'acide aminé (étiquette) de 3 à 5 AAs est déterminée. En combinant la masse  $m$  du peptide, les masses  $m_1$  et  $m_2$  des extrémités du peptide et l'étiquette, et en recherchant ces informations dans les bases de données de protéines, il est possible d'identifier le peptide. Plusieurs étiquettes peuvent être utilisées pour le même peptide afin d'augmenter la probabilité d'une identification correcte.

## 3.2 Identification de microorganismes

Il existe aussi des logiciels spécialisés dans l'identification de microorganismes par la reconnaissance de leur empreinte moléculaire. En effet, les spectres MALDI-TOF des microorganismes fournissent un profil protéique qui leur est propre permettant ainsi de réaliser des identifications. Les trois principaux fournisseurs sont Bruker avec le Biotyper, Biomérieux avec VITEK® MS et Andromas.

### 3.2.1 Un peu d'histoire

Les premières identifications de bactéries par spectrométrie de masse datent de 1975 par couplage pyrolyse/spectrométrie de masse (Anhalt and Fenselau, 1975). Krishnamurthy et al. (Krishnamurthy and Ross, 1996; Krishnamurthy et al., 1996) sont les premiers à utiliser le MALDI-TOF pour différencier des bactéries pathogènes de bactéries non-pathogènes en comparant leurs spectres à la suite d'une lyse mécanique ou chimique. Il est déjà précisé qu'avec seulement une calibration externe, c'est-à-dire une calibration non spécifique à l'échantillon mais appliquée à l'ensemble des échantillons, la précision sur les masses est limitée (0,2 % de déviation, soit 2 000 ppm).

L'équipe de Haag et al. a étudié la différence de spectres obtenus par ces cellules entières et par des cellules lysées de *Haemophilus* (Haag et al., 1998). L'analyse MALDI-TOF de cellules entières et du surnageant obtenu après rinçage des cellules avec un mélange eau/acétonitrile est identique avec l'observation de pics intenses autour de valeurs de  $m/z$  de 10 000. En revanche, la comparaison des spectres des cellules entières et des cellules lysées montre des spectres très différents avec majoritairement des pics aux faibles masses, de  $m/z$  entre 4 000 et 6 000 et une disparition dans le bruit de fond des pics de valeurs de  $m/z$  de 10 000 pour les cellules lysées. Le travail de Saenz et al. a consisté à étudier les résultats obtenus par MALDI-TOF pour 10 séries de cultures sur une durée de trois mois pour deux types d'espèces (*Escherichia coli* et *Bacillus atrophaeus*) (Saenz et al., 1999). Il en résulte que l'étude MALDI-TOF de bactéries est très reproductible au cours du temps si les mêmes méthodes de préparation sont utilisées. Il apparaît donc qu'il est possible d'attribuer des spectres de références spécifiques à chaque microorganisme et donc d'identifier des microorganismes rapidement à partir de ces spectres. Ceci montre aussi l'importance du mode de préparation des échantillons pour obtenir des spectres répétables pour l'identification. Toutefois, pour comparer le spectre d'un échantillon inconnu à une bibliothèque de spectres, il est nécessaire de créer cette bibliothèque et donc posséder les souches des microorganismes et que ceux-ci soient cultivables. Afin d'automatiser les identifications et de donner une valeur aux résultats obtenus, plusieurs types d'algorithmes ont été mis en place.



### 3.2.2 Exemples de méthodes d'identification de microorganismes

#### a. Comparaison de sections de spectres

Arnold et Reilly ont comparé les scores de corrélation croisée entre des spectres entiers et entre des spectres divisés en 13 intervalles de 500 Da entre 3,5 et 10 kDa (Arnold and Reilly, 1998). La corrélation croisée est la mesure de similitude entre deux signaux. Plus les signaux sont semblable plus le score est proche de 1 et inversement, plus les signaux diffèrent, plus le score est proche de 0. Pour chaque intervalle la corrélation croisée entre les spectres à comparer est calculée. Les corrélations croisées de chaque intervalle sont normalisées par la moyenne des valeurs des corrélations croisées. La corrélation composite des deux spectres est calculée comme le produit des valeurs de corrélation croisées normalisées de chaque intervalle. Cette méthode repose principalement sur la reproductibilité des spectres puisque l'ensemble du spectre est pris en compte et comparé avec le spectre de référence contrairement aux autres méthodes détaillées ci-dessous. Plus le nombre d'intervalles augmente plus la méthode est sélective et plus il est important que les spectres soient reproductibles. L'avantage de cette méthode réside dans le fait que la distinction entre chaque souche est précise puisque chaque détail du spectre est pris en compte (distinction de 25 souches de *E. coli*). La séparation en différents intervalles amplifie les différences entre les scores des spectres en comparaison avec la corrélation croisée de deux spectres entiers. Ceci permet de bien différencier deux souches différentes : pour les spectres divisés en intervalles le score est de  $2,260 \times 10^{-12}$  ce qui montre bien que les spectres présentent de nombreuses différences alors que le score de corrélation croisée pour les mêmes spectres pris entiers est de 0,4572, ce qui rend la différence entre les spectres moins évidente.

Cependant, les méthodes d'identification de microorganismes par comparaison de spectres n'incluent pas la notion de dynamique des protéines. En effet, l'expression des protéines est tout particulièrement influencée par les différents stimuli externes liés à l'environnement de la bactérie. De ce fait, selon les conditions de croissance les profils de bactéries, qui sont l'image du contenu en protéines de l'échantillon, seront différents pour un même microorganisme. Ce qui signifie que pour qu'il n'y ait pas d'erreurs d'identification, la banque de spectres devrait contenir les données obtenues pour un même microorganisme à partir de différentes conditions de culture.

#### b. Recherche dans des bases de données

Comme les recherches par comparaison de spectres sont extrêmement dépendantes des conditions et temps de culture mais aussi des conditions d'analyses, une autre possibilité explorée par Demirev *et al.* et Pineda *et al.*, a été de rechercher les masses des protéines sur des bases de données telles que Swiss-Prot (Demirev *et al.*, 1999; Pineda *et al.*, 2000). L'intérêt de cette méthode est que le résultat ne dépend plus des conditions de culture, de la reproductibilité des spectres, des contaminants ou de la gamme de masse étudiée. En effet, quelles que soient les protéines exprimées par le microorganisme ou la machine utilisée pour enregistrer les spectres, les microorganismes étaient a priori identifiables (dans la mesure où leur nombre dans la base est restreint et où leur génome est totalement



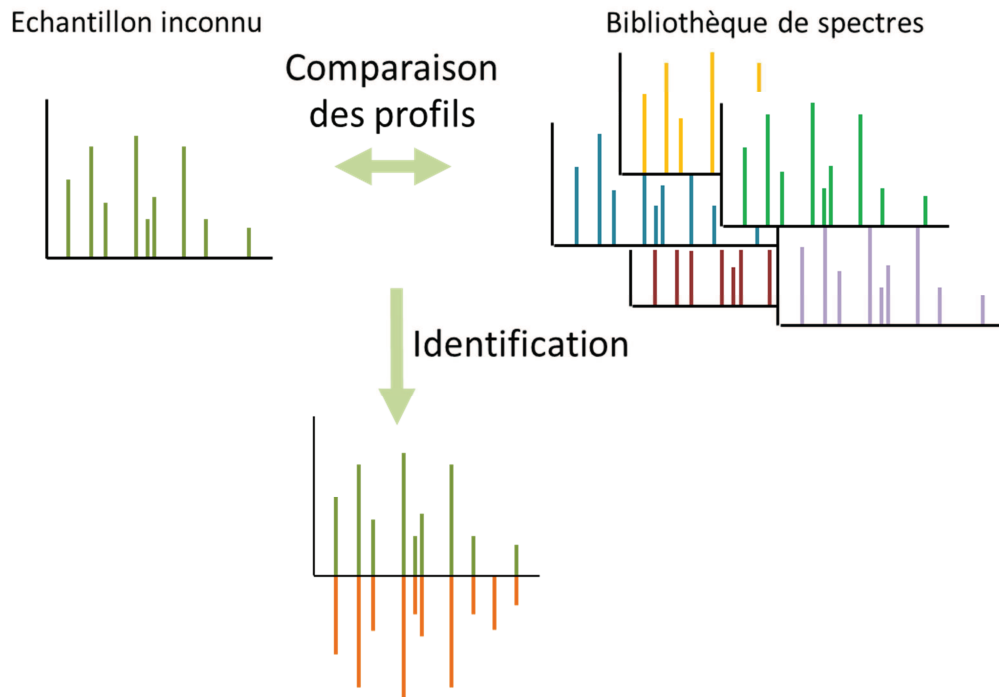
séquencé). Désormais, la grande quantité de protéines contenues dans les bases de données rend aléatoire l'attribution d'une masse à une protéine d'un organisme et un grand nombre d'erreurs d'identification peut se produire. Cette technique n'est envisageable que sur des bases de données restreintes. En 2000, au moment de la publication de Pineda *et al.*, Swiss-Prot ne contenait que 219 microorganismes avec plus de 10 séquences de protéines. Aujourd'hui dans la version 2014-06 de UniProtKB, composé des bases Swiss-Prot dont les séquences sont validées manuellement et TrEMBL dont les séquences sont annotés automatiquement à partir des génomes, Swiss-Prot possède 338 085 séquences appartenant à 1 738 bactéries. Parmi ces bactéries, 897 ont plus de 10 séquences (au total 337 118 séquences pour ces 897 bactéries contre 17 652 séquences pour 219 microorganismes en 2000). Actuellement la base TrEMBL possède 65 532 440 séquences appartenant à des bactéries. La taille des bases de données croissant de manière exponentielle, ce type de méthode est donc difficilement applicable. Aujourd'hui, la recherche d'une protéine de 10 000 Da sur la base entière de NCBI-nr avec une erreur de 100 ppm fournit 7 665 réponses dont 6 800 réponses avec une erreur de 10 ppm (recherche effectuée sur TagIdent, plateforme ExpASY, <http://web.expasy.org/tagident/>). De plus, cette méthode ne tient pas en compte des modifications post-traductionnelles possibles (perte de méthionine, acétylation, etc.) ce qui favorise d'autant plus la probabilité d'avoir une erreur d'attribution. En revanche, l'établissement d'un facteur de pertinence du résultat obtenu par l'algorithme tient compte de la taille des protéomes associés à chaque organisme. En effet, certains organismes en cours de séquençage ne possèdent qu'une dizaine de protéines et des bactéries comme *E. coli* en possèdent environ 2 000 entre 4 et 20 kDa. Ainsi, quelques protéines identifiées chez un organisme seront très pertinentes si son protéome est de taille réduite, mais le même nombre chez *E. coli* ne donnera pas pour autant un fort taux de pertinence. Pour améliorer cette méthode il faudrait i) tenir compte des modifications post-traductionnelles et des charges multiples ; ii) améliorer les techniques de mesures pour détecter plus de protéines et avec plus de précision ; iii) réduire la taille des bases de données à ce qui serait observable (Pineda *et al.*, 2000). L'identification en tenant compte de la perte possible de méthionine a été abordée dans une nouvelle publication en 2001 par la même équipe (Demirev *et al.*, 2001). Cette fois-ci c'est l'identification de *H. pylori* qui est réalisée contre 18 organismes possédant au moins 200 protéines dans l'intervalle de 4 à 20 kDa. L'identification en tenant compte de la possibilité de perte de méthionine a permis de gagner un ordre de grandeur sur le facteur de pertinence (passant de 0,036 à 0,002). Devant la quantité de modifications post-traductionnelles possibles et devant la taille croissante des bases de données, ce type de méthode n'a pas d'avenir pour l'identification par MALDI-TOF/MS.

### c. Comparaison avec des spectres de référence composés de biomarqueurs

Dans la publication de Jarman et *al.* (Jarman et al., 1999) plutôt que de comparer les spectres dans leur globalité, ce sont des valeurs de  $m/z$  qui sont extraites des spectres MALDI-TOF et qui font office de biomarqueurs. Les spectres de référence auxquels seront comparés les spectres des échantillons inconnus à identifier, sont construits comme suit : une table regroupe les rapports  $m/z$  et leurs intensités pour chaque signal ainsi que leur fréquence de présence dans les spectres ayant servi à constituer le spectre de référence. Ce dernier élément tient compte de la possibilité que les spectres varient par la différence d'expression des protéines ou la non-détection d'un pic. Chaque biomarqueur est modélisé par une boîte représentant l'intervalle de 95 % où se trouvent les pics ayant servi à constituer les références (en largeur et en hauteur c'est-à-dire selon leur rapport  $m/z$  et leur intensité). Ici, un intervalle d'intensité est considéré mais la préoccupation statistique première est la valeur  $m/z$  du pic. Ensuite la détection d'un microorganisme est basée sur la présence dans le spectre de l'échantillon inconnu de ces biomarqueurs. Les pics à comparer aux spectres de références sont sélectionnés à partir d'un test de Student. Cette méthode permet de comparer les spectres uniquement sur des pics pertinents en s'affranchissant de pics pouvant être dus à la matrice ou à des contaminants et tient compte de la variabilité spectrale en ajoutant un facteur de fréquence au moment de la création des spectres de référence. En revanche, la conservation des protéines au sein d'une espèce ou d'une sous-espèce peut entraîner des confusions en identifiant une souche proche de celle présente dans l'échantillon testé. De plus, des erreurs de calibration peuvent entraîner un décalage de masses rendant toute identification impossible.

### 3.3 Systèmes intégrés d'identification de microorganismes

Aujourd'hui, des méthodes d'identification à partir des spectres MALDI-TOF sont automatisées et des bibliothèques de spectres existent. Ces méthodes sont couramment utilisées en routine dans les CHU pour identifier des bactéries, des champignons ou des moisissures (Blondiaux et al., 2010; Cassagne et al., 2011). En revanche, les recherches de masses de protéines entières dans les bases de données ont peu évolué. Nous allons maintenant détailler les systèmes existants. Ils se basent sur la corrélation entre le spectre d'un échantillon et une librairie de spectres de référence comme le schématise la **Figure 11**. Ce sont des systèmes intégrés composés d'un spectromètre de masse, du logiciel d'identification et d'une librairie de spectres. L'utilisateur peut compléter les bases de données par ses propres spectres de références.



**Figure 11 :** Comparaison de profils protéiques pour l'identification de microorganismes

### 3.3.1 Biotyper (Bruker)

Les spectres de référence sont générés à partir des pics spécifiques contenus dans au moins 20 spectres obtenus avec différents paramètres de culture. La comparaison de spectres inconnus avec les spectres de référence se base à la fois sur le rapport  $m/z$  des pics, la fréquence et l'intensité. Il est possible de recalibrer les spectres après acquisition et d'ajouter une fenêtre d'erreur. Au moins  $10^7$  bactéries sont nécessaires pour une identification correcte sur un appareil de type Microflex (Bruker) ou Ultraflex I. Les résultats sont fournis sous la forme d'une liste de microorganismes identifiés en fonction du score obtenu. Si le score est compris entre 3 et 2,33, l'identification est du niveau espèce. Entre 2,32 et 2, l'identification est du niveau genre et probablement du niveau espèce. Entre 1,99 et 1,7, l'identification est probablement du niveau genre et un score inférieur à 1,69 ne donne pas d'identification (Sauer et al., 2008). Actuellement, plus de 5 600 spectres de références sont contenus dans la librairie du Biotyper.

### 3.3.2 Andromas (Andromas SAS)

L'équipe de Xavier Nassif (hôpital Necker) a développé ce système pour les microbiologistes. Il est constitué d'un MALDI-TOF de paillasse construit pour cette application spécifique, d'une base de données bactérienne et fongique et d'un logiciel d'identification. Les spectres de référence sont créés à partir de 10 spectres d'un microorganisme et seuls les pics présents dans tous les spectres, et dont l'intensité relative est supérieure à 0,1, sont retenus pour créer le spectre de référence. A partir d'un spectre inconnu, tous les pics dont l'intensité relative est supérieure à 0,1 sont extraits et comparés avec les spectres de référence en tenant en compte de possibles variations de  $m/z$  de  $\pm 10$ . Le résultat est fourni sous la forme d'un pourcentage de pics communs calculé ainsi :  $100 \times (\text{nombre de pics}$

commun/nombre de pics spécifiques dans le spectre de référence). Si le pourcentage de pic commun est supérieur à 70 % et que la différence entre la première identification et la seconde est d'au moins 10 % alors l'identification est valable (Alanio et al., 2011; Carbonnelle et al., 2007; Lacroix et al., 2014).

### 3.3.3 VITEK® MS (bioMérieux)

Le VITEK® MS de bioMérieux (<http://www.vitekms.com/>) est né de l'association de deux partenaires : Shimadzu, le constructeur du MALDI-TOF, et Anagnostech qui a mis au point la base de données Aramis® et l'algorithme d'identification. Deux bases de données sont disponibles. La première base de données est la base Aramis®, dite « ouverte », qui contient les spectres de 1 210 microorganismes. L'algorithme d'identification est celui créé par Anagnostech. La seconde base de données, dite « fermée », est composée de 755 microorganismes. Elle est mise à jour régulièrement et son algorithme, Advance Spectrum Classifier (ASC), a été mis au point par bioMérieux. Ce deuxième système a reçu l'approbation de l'Agence américaine des produits alimentaires et médicamenteux (Food and Drug Administration, FDA) pour la commercialisation aux Etats-Unis. Pour créer la bibliothèque de spectres, chaque microorganisme de référence est analysé avec différents modes de cultures et par différents utilisateurs sur différents instruments pour en moyenne 10 isolats par espèce. Le premier algorithme se base sur la méthode dite des « SuperSpectres ». Des spectres « moyens » sont créés pour les familles, les genres ou les espèces en sélectionnant parmi les spectres de référence entre 20 et 30 pics spécifiques à la famille au genre ou à l'espèce. Les spectres des échantillons inconnus sont d'abord comparés aux SuperSpectres puis aux spectres de référence si la première comparaison n'a pas pu fournir de résultats. L'algorithme ASC est une extension de SuperSpectres avec en plus la notion de présence ou d'absence des pics (van Belkum et al., 2012). Les spectres acquis sur la gamme de  $m/z$  variant de 3 000 à 17 000 sont divisés en 1 300 intervalles et à chaque intervalle est attribué un poids qui définit la spécificité, ou non, de la présence, ou de l'absence, d'un pic dans cet intervalle en fonction des différents taxons. Ensuite pour chaque espèce la distribution des poids par intervalle est comparée à celle de l'ensemble des autres espèces. Une fonction mathématique est dérivée de la comparaison des deux distributions pour chaque espèce. La somme des poids pour un spectre d'une espèce inconnue peut alors être prise en charge par cette fonction pour donner un taux de confiance (pourcentage) sur l'attribution d'un spectre inconnu à une espèce. Un score au-delà de 60 % montre que la souche inconnue fait partie de la distribution de score de l'espèce et le résultat est de confiance. Si plus de 4 espèces sont dans le rapport final ou s'il n'y en a aucune, le résultat « pas d'identification » est retourné (Rychert et al., 2013).

## 3.4 Publications des résultats d'identification

Comme nous avons pu le voir la spectrométrie de masse et l'interprétation des données fournissent une grande variété et une grande quantité d'informations. Il a donc été nécessaire d'établir des règles strictes concernant la validation des données d'identification ou de quantification ainsi que leur restitution dans le cadre de publications. D'importants

journaux comme Molecular and Cellular Proteomics (MCP) (Carr et al., 2004) et des organismes internationaux comme HUPO (Human Proteome Organization) avec le MIAPE (Minimal Information About a Proteomics Experiment) ont mis en place des « règles de bonne conduite » pour le traitement des données de protéomique et la soumission de celles-ci (Taylor et al., 2007). Le MIAPE MS décrit les données qui doivent être fournies à propos de l'acquisition et du traitement des données (Taylor et al., 2008). Ces recommandations sont mises à jour au maximum une fois par an, assurant ainsi une stabilité et une continuité des données transmises par les auteurs aux journaux. Dans ces recommandations il n'y a pas d'informations sur la manière dont les données doivent être traitées telles que des valeurs seuils ou des protocoles à suivre. Il s'agit, en revanche, d'avoir les informations nécessaires pour comprendre la manière dont les données sont obtenues puis traitées afin de pouvoir éventuellement reproduire ou retraiter ces informations. Parmi les informations à fournir il y a notamment les scores attribués par les logiciels d'identification aux peptides ou aux protéines, leur e-value et le taux de faux positifs.

#### **4 Méthodes d'enrichissement de bactérie pour une identification par spectrométrie de masse MALDI-TOF.**

Nous allons détailler les différentes techniques d'enrichissement existantes et les travaux qui ont été menés ces dernières années pour enrichir des échantillons complexes dans le but d'identifier des microorganismes à l'aide de la spectrométrie de masse.

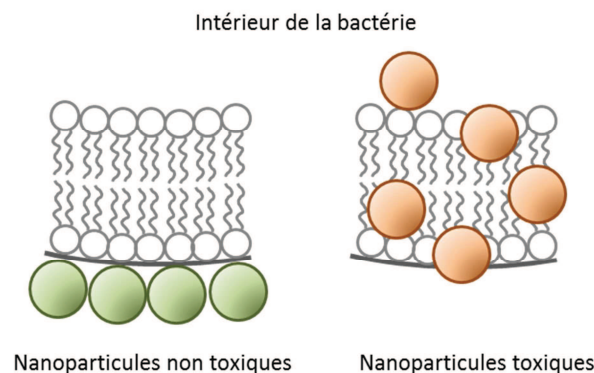
Pour abaisser les limites de détection des bactéries, leur enrichissement est nécessaire. Il peut être spécifique d'une espèce d'un microorganisme (Madonna et al., 2001; Madonna et al., 2003; Ochoa and Harrington, 2005) ou concerner l'ensemble des microorganismes présents dans l'échantillon (Chen et al., 2008; Gopal et al., 2011; Guo et al., 2009; Li et al., 2009; Li et al., 2010; Lin et al., 2005).

##### **4.1 L'immunocapture magnétique**

L'immunocapture à l'aide de billes magnétiques sert à concentrer des échantillons en protéines ou bactéries. Différents types de surfaces fonctionnalisées sont disponibles dans le commerce. Certaines, telles que les paratoluène sulfonyl (= tosyl), permettent de greffer de manière covalente des anticorps à la surface des billes tels que des anti-peptides, des anti-protéines ou des anti-bactéries entières (Schlosser et al., 2007). Les billes magnétiques anti-*Salmonella* (Dynabeads® anti-*Salmonella*, Life Technologies) ont permis d'atteindre sur un MALDI-TOF de type Voyager-DE STR (AB Applied Biosystems) une limite de détection de  $10^7$  CFU/ml - concentration initiale de l'échantillon - pour des *Salmonella* entières (Madonna et al., 2001; Ochoa and Harrington, 2005). La détection indirecte des bactéries par la capture de bactériophage permet d'atteindre une limite de détection de  $5 \times 10^4$  CFU/ml (Madonna et al., 2003).

## 4.2 Les nanoparticules (NPs)

Plus récemment la technologie des nanoparticules a fait son apparition permettant l'enrichissement de manière non spécifique. Les nanoparticules peuvent soit permettre de séparer les bactéries du reste de l'échantillon en utilisant une séparation magnétique (Lin et al., 2005), soit permettre l'ionisation ciblée, grâce aux nanoparticules liées aux bactéries, des protéines de la bactérie cible (Wu et al., 2012). Les nanoparticules sont des sphères d'environ 11 nm de diamètre contre 2,8  $\mu\text{m}$  de diamètre pour les billes magnétiques. Les nanoparticules de  $\text{Fe}_3\text{O}_4$ , Au, Pb, Pt, NiO, Ag,  $\text{TiO}_2$ , ZnO, ont une affinité naturelle avec les bactéries, mêmes lorsqu'elles ne sont pas modifiées. Contrairement aux autres nanoparticules qui interagissent uniquement avec l'extérieur des membranes cellulaires ( $\text{Fe}_3\text{O}_4$ , Au, Pb, Pt, NiO), les nanoparticules de Ag,  $\text{TiO}_2$ , ZnO sont toxiques pour les bactéries car elles pénètrent la membrane cellulaire des bactéries comme le montre la **Figure 12**. Pour celles-ci, il existe une concentration critique à partir de laquelle la résolution et la sensibilité des spectres diminuent (Wu et al., 2012). Il est possible de greffer les nanoparticules, avec de la vancomycine par exemple, favorisant ainsi les interactions avec les bactéries gram-positives (Lin et al., 2005).



**Figure 12** : Interaction des nanoparticules avec la bactérie selon leur toxicité. Les nanoparticules non toxiques (schéma de gauche) se fixent à la surface de la bactérie. Les nanoparticules toxiques (schéma de droite) ont des propriétés bactéricides et oxydent la membrane avant de pénétrer à l'intérieur de la bactérie (Wu et al., 2012).

Gopal et al. (Gopal et al., 2011) obtiennent une limite de détection de bactéries entières de  $4,6 \times 10^2$  CFU/ml (*Staphylococcus aureus*). Cependant, les spectres obtenus à partir de prélèvements buccaux et nasaux montrent un nombre important de pics pour des  $m/z$  variant de 2 000 à 11 000 qui ne sont pas tous attribuables à *Staphylococcus aureus*.

## 4.3 Les surfaces fonctionnalisées – SELDI

Plus rares mais également utilisées, il existe des surfaces fonctionnalisées. Les surfaces peuvent être greffées avec des anticorps (Brauer et al., 2010; Johnson et al., 2009; Kim et al., 2012; Liang et al., 1998; Patrie and Mrksich, 2007), des peptides (Johnson et al., 2009), des lipopolysaccharide (LPS) (Thirumalapura et al., 2005) et des lectines (Afonso and Fenselau, 2003). Plusieurs techniques sont évoquées dans la littérature : le greffage sur monocouche d'or (Johnson et al., 2009; Kim et al., 2012), l'assemblage de monocouches (Brauer et al., 2010; Patrie and Mrksich, 2007; Stollowitz and Stephan, 2009) et l'utilisation de

nitrocellulose (Liang et al., 1998; Thirumalapura et al., 2005). Le contrôle de l'hydrophobicité de la surface accroît les capacités d'enrichissement de celle-ci (Brauer et al., 2010; Kim et al., 2012). Le bémol de cette technique, qui présente pourtant l'avantage d'être très rapide puisque le dépôt sur la cible est direct, est la faible quantité d'anticorps pouvant être greffée. En effet, seulement 70 fmol/mm<sup>2</sup> (Johnsson et al., 1991) soit 700 fmol pour un spot de 10 mm<sup>2</sup> pour les surfaces fonctionnalisées contre 3 µmol/ml pour les Dynabeads® M-280 Tosylactivated d'après les données du constructeur (Life Technologies), soit pour un volume de billes de 100 µl, environ 300 nanomoles d'anticorps fixés. La publication traitant d'enrichissement de bactéries par surfaces fonctionnalisées avec des anticorps ne mentionne pas le nombre de bactéries fixées par cette technique (Afonso and Fenselau, 2003). L'utilisation de spectromètres de masse suffisamment sensibles pour détecter une petite quantité de bactéries pourrait résoudre le problème de la limite de détection qu'entraîne cette méthode.



## 5 Objectifs du travail de thèse

### 5.1 Le projet Franco-Allemand GEFREASE

Au cours de ces dernières années, les nations occidentales ont renforcé leurs réseaux et moyens pour faire face aux menaces bioterroristes. La mise en place d'outils pour l'identification des menaces biologiques a permis l'émergence d'un certain nombre de biosenseurs basés sur diverses techniques (couplage moléculaire, senseurs d'affinité, mesures d'activité...). Au vu des technologies existantes, les experts en France et en Allemagne ont identifié différentes limitations qui restreignent l'usage des solutions technologiques disponibles. Une des limites principales est que, malgré de véritables savoir-faire académiques et industriels, il n'existe pas d'effort transnational pour développer, valider et harmoniser la panoplie de technologies existantes. De ce contexte est née la proposition baptisée GEFREASE (GERman FRENch Equipment for Analysis and Surveillance of biothreats in the Environnement, Equipements franco-allemands pour l'analyse et la surveillance des menaces biologiques dans l'environnement) : deux groupes académiques et deux groupes industriels chacun en France et en Allemagne avec un fort passé scientifique et technologique en biodéfense, se sont associés pour rassembler leurs compétences afin de construire une plateforme de biodétection pour les menaces environnementales liées à des actions de malveillance bioterroristes. Le premier partenaire est le Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA, France), dont une composante, la Direction des Sciences du Vivant est leader national pour les programmes de biodéfense ; le second partenaire académique est le Robert Koch Institut (RKI, Allemagne), l'organisme central allemand pour le contrôle des maladies infectieuses et également centre de référence dans le domaine du secteur public de la santé en Allemagne. Deux partenaires industriels leaders dans le domaine NRBC-E font partie du projet : le premier est Bruker (Leipzig, Allemagne) qui possède une longue expérience dans le développement d'outils de spectrométrie de masse et d'outils portables de mesure NRBC-E et Bertin Technologies (France) qui développe des systèmes de collecte et de traitement d'échantillons à des fins d'analyses biologiques. Le CEA est responsable de la coordination et de la direction du projet, divisé en 6 tâches. Sur la base de savoir-faire existants, les quatre partenaires ont pour objectif de développer une approche de diagnostic intégrée combinant un système de détection sur site (première intervention et/ou détection provisoire basée sur des techniques immunoanalytiques) avec une approche de confirmation de laboratoire basée sur des analyses par spectrométrie de masse. Les cibles seront parmi les agents les plus connus du risque biologique, à savoir : des toxines (ricine, toxines botuliniques de type A, B, C, D, E et F, entérotoxines A et B, abrine) et des microorganismes bactériens et viraux (*Francisella tularensis*, *Bacillus anthracis*, *Yersinia pestis*, *Vibrio cholerae* et pox virus). Notre but est de construire une plateforme pour détecter et identifier ces différents agents dans divers milieux environnementaux (air, nappes aquifères, eau de consommation, boissons...). Le but ultime de ce consortium est i) de fédérer au niveau européen les forces franco-allemandes existantes, ii) de développer des modèles et des outils capables de répondre à des besoins de Sécurité Civile pour la détection



précise des menaces biologiques et iii) de créer un réseau de force de proposition pour des programmes transnationaux et européens en technologies de Sécurité Civile. Ce projet a obtenu un financement de l'Agence Nationale de la Recherche (ANR) qui a démarré en mai 2012.

## 5.2 Les objectifs du projet de thèse entrepris

Le projet de thèse s'inscrit dans le cadre du projet GEFREASE et concerne la partie confirmation par des analyses de spectrométrie de masse. L'objectif du travail de thèse est d'identifier des marqueurs antigéniques ou peptidiques permettant la caractérisation des microorganismes pathogènes et de mettre en place des approches ciblées pour la préparation d'échantillon avec des méthodes immunologiques en amont de la spectrométrie de masse MALDI-TOF. Ce sujet s'articule sur la compétence reconnue (microbiologie, spectrométrie de masse et immunologie analytique) de plusieurs équipes du SBTN.

Une première partie du travail a été consacrée à l'étude du protéome de la bactérie modèle *Francisella tularensis* subsp. *holarctica* LVS, microorganisme du programme GEFREASE, afin d'identifier les protéines et les peptides les plus abondants donnant un signal intense par spectrométrie de masse MALDI-TOF, dans le but de pouvoir conclure sur la virulence d'une souche par la connaissance de la sous espèce à laquelle elle appartient. Un article sur le travail réalisé sur le choix des protéines cibles pour l'identification par spectrométrie de masse des bactéries du genre *Francisella* a été accepté dans le journal *Analytical Chemistry*.

La seconde partie du travail a été consacrée à l'enrichissement de la bactérie modèle par immunocapture magnétique dans le cadre de la préparation d'échantillon avant l'analyse par spectrométrie de masse MALDI-TOF pour confirmer les résultats obtenus sur le terrain. Certaines contraintes techniques dues à l'utilisation éventuelle d'un laboratoire mobile ont dû être prises en compte telles que, l'absence de centrifugeuse ou la nécessité de protocoles simples pouvant être réalisés par des primo-intervenants (pompiers par exemple). L'objectif est de pouvoir concentrer la bactérie modèle dans des échantillons complexes où elle est largement minoritaire et dans des échantillons de diverses matrices alimentaires telles que du lait ou de l'eau minérale.

# Résultats



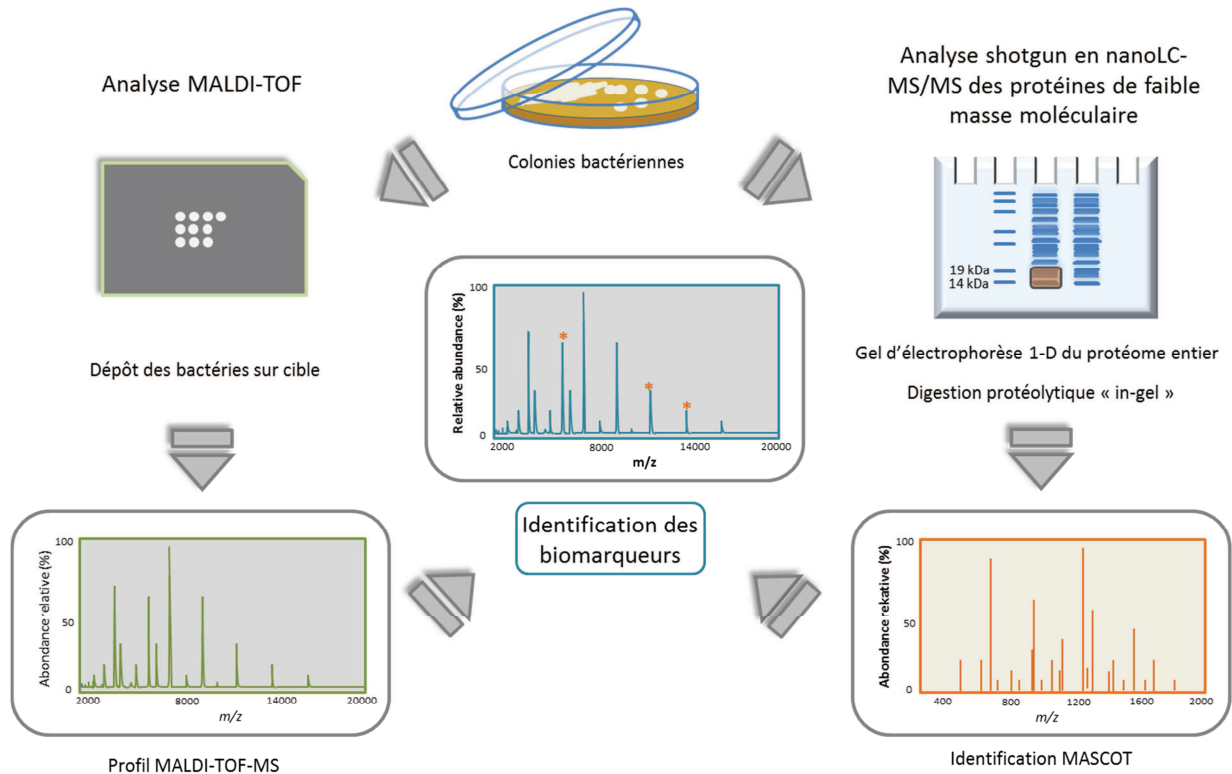
# 1<sup>ère</sup> partie : Identification de *Francisella tularensis* jusqu'au niveau sous-espèce par la détection de biomarqueurs spécifiques

## 1 Contexte et stratégie proposée

L'identification des microorganismes par spectrométrie de masse de type MALDI-TOF actuellement couramment employée dans les laboratoires d'analyse microbiologique est principalement basée sur la reconnaissance de profil protéique. En effet, les protéines des bactéries présentent, pour une condition physiologique donnée, lors d'une analyse MALDI-TOF, un profil de masses particulier qui leur est propre à la fois en termes de rapport masse sur charge ( $m/z$ ) et d'intensité. L'identification se fait par comparaison du spectre enregistré à des spectres de référence acquis à partir de microorganismes d'origine connue. Il n'est pas nécessaire de connaître les protéines à l'origine des signaux du spectre MALDI-TOF pour réaliser ces identifications puisque c'est un processus de corrélation entre les profils protéiques des spectres qui est mis en jeu. Les protéines qui forment ces profils sont majoritairement des protéines ribosomales qui sont très conservées au sein de chaque espèce. Elles ont la particularité d'être de petite masse moléculaire et basique. Il n'y a donc que quelques protéines qui différencient les microorganismes au sein d'une même espèce. Pouvoir différencier les microorganismes selon leur sous-espèce est un véritable enjeu notamment lorsque certaines sous-espèces ou souches sont plus ou moins pathogènes. Le but de cette partie est d'identifier les protéines qui permettent de différencier les microorganismes virulents des microorganismes non pathogènes. Il s'agit ainsi d'identifier des biomarqueurs de la présence de bactéries virulentes dans un échantillon. Pour cela nous avons choisi comme microorganisme modèle *Francisella*. Cette espèce est composée de plusieurs espèces de virulences différentes. Les trois sous-espèces de *F. tularensis*: *tularensis*, *holarctica*, et *mediasiatica*, sont de virulence décroissante pour l'homme, avec un taux de mortalité de l'ordre de 30 % pour une exposition à *F. tularensis* subsp. *tularensis*. Les espèces *novicida* et *philomiragia*, bien que pathogènes chez l'animal, ne sont pas pathogènes chez l'homme. Le laboratoire dispose de la souche vaccinale de *Francisella*, *Francisella tularensis* subsp. *holarctica* LVS et de *Francisella philomiragia* ATCC 25015, qui se manipulent en laboratoire de biosécurité de niveau 2 et *Francisella tularensis* subsp. *tularensis* CCUG 2112 qui se manipule en laboratoire de biosécurité de niveau 3.

La stratégie suivie pour l'identification des biomarqueurs est décrite par la **Figure 13**. Après inactivation de *Francisella tularensis* subsp. *holarctica* LVS, son protéome sera étudié par spectrométrie de masse MALDI-TOF et nanoLC-MS/MS. L'identification des protéines les plus intenses sera suivie de l'étude de celles-ci sur l'ensemble des souches de l'espèce de

*Francisella* dont les génomes sont séquencés ou en cours de séquençage. Ceci permettra de choisir des biomarqueurs répondant au critère défini précédemment, c'est à dire capables de différencier facilement les échantillons contenant des souches de *Francisella* virulentes des échantillons contenant des souches de *Francisella* non virulentes.



**Figure 13** : Stratégie d'identification des biomarqueurs utilisant MALDI-TOF et nanoLC-MS/MS. Tiré de (Christie-Oleza et al., 2013). Les colonies bactériennes sont étudiées à la fois par analyse MALDI-TOF et par analyse shotgun. Les spectres MALDI-TOF permettent d'identifier les valeurs  $m/z$  des pics les plus intenses et l'analyse shotgun permet d'identifier les protéines qui sont exprimées par la bactérie. En recoupant les informations obtenues par les deux types d'analyses il est possible d'identifier les protéines donc les pics sont intenses sur les spectres MALDI-TOF. Il est alors possible d'identifier des protéines spécifiques à chaque microorganismes sur les spectres MALDI-TOF.

## 2 Caractérisation des protéines les plus abondantes par analyse protéomique

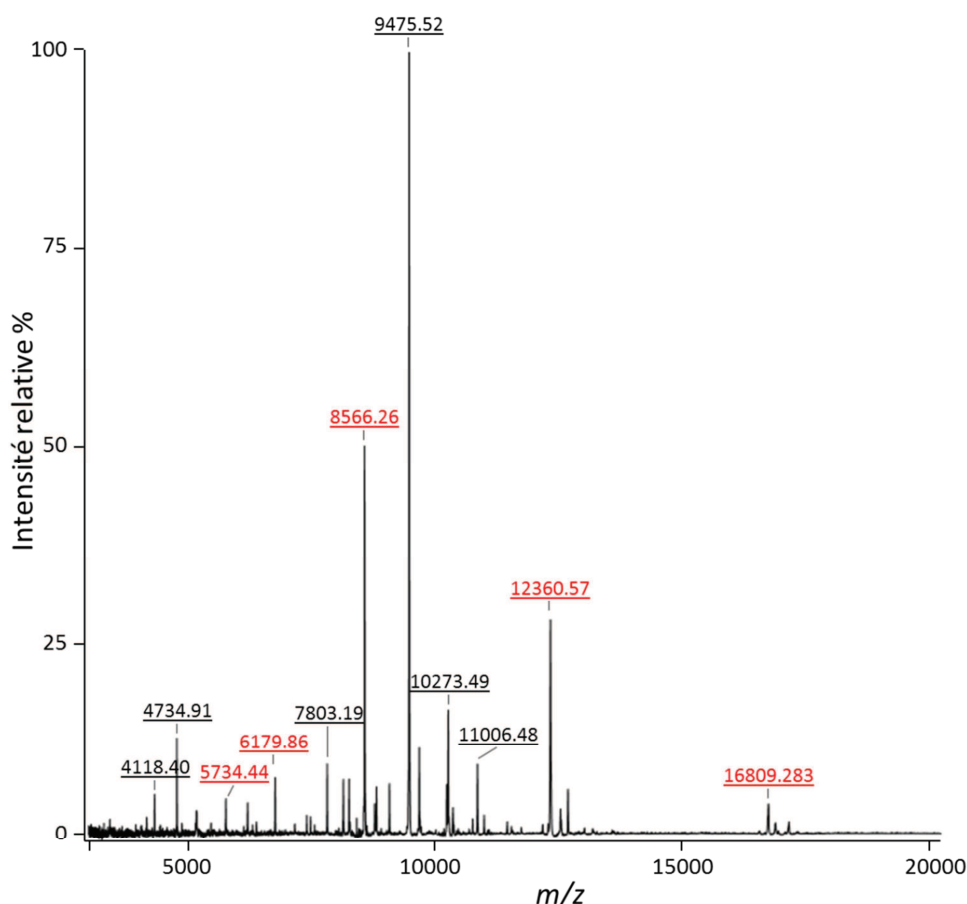
### 2.1 Détection MALDI-TOF après inactivation au TFA

*Francisella tularensis* subsp. *holarctica* LVS est une bactérie qui nécessite un milieu enrichi pour son développement. Les bactéries sont donc cultivées sur des boîtes de géloses de type « chocolat », appelées ainsi du fait de leur couleur brune donnée par le sang cuit qui entre dans la composition du milieu. Après ensemencement et croissance à 37°C de ces bactéries, les colonies sont ensuite mises en culture dans un milieu liquide BHI (Brain Heart Infusion) enrichi en PolyViteX (PVX, 1 %).

Afin de pouvoir travailler sur ces bactéries pathogènes en toute sécurité (manipulation en laboratoire de biosécurité de niveau 2), un protocole d'inactivation à l'acide

trifluoroacétique (TFA) basé sur le travail de Lasch *et al.* a été testé et validé (Lasch et al., 2008). La forte concentration d'acide dans ce protocole (80 %) lors de la première incubation assure une lyse des cellules. L'étape finale de centrifugation élimine les débris cellulaires et les bactéries qui auraient résisté à la lyse afin de ne conserver que le surnageant qui contient les protéines. La validation de l'inactivation a été réalisée en déposant 100 µl de ce surnageant sur une boîte de gélose chocolat + PVX. La validation de l'inactivation a été obtenue puisque aucune colonie n'a poussé durant les 6 jours d'incubation à 37°C. Les échantillons obtenus par cette méthode ont pu être soit déposés sur cible MALDI, soit protéolysés pour être analysés en nanoLC-MS/MS.

Une étape de mise au point des conditions de dépôt sur cible MALDI a été nécessaire pour avoir la meilleure calibration possible. Tout d'abord l'acide sinapinique (SA) comme matrice est le choix le plus judicieux puisqu'il est recommandée pour l'étude en MALDI de protéines de masses supérieures à 3 kDa, alors que l'acide  $\alpha$ -cyano-4-hydroxycinnamique (CHCA) est plutôt recommandé pour l'étude de peptides. Une calibration externe a été réalisée en déposant entre les spots d'échantillons un calibrant, le « protein calibration standard I » de chez Bruker. Une seconde calibration, interne, a été effectuée en ajoutant directement le calibrant à l'échantillon. Ce volume de calibrant doit être adapté en fonction de la quantité de protéines contenu dans l'échantillon à analyser. Sans cette optimisation, deux phénomènes peuvent se produire : si le calibrant est trop intense, il atténue le signal de l'échantillon (bruit de fond), et inversement, si le signal du calibrant est trop faible, il est impossible d'effectuer la calibration interne. L'utilisation de la calibration interne est recommandée pour obtenir les masses les plus précises possibles. La **Figure 14** présente un spectre MALDI-TOF représentatif du protéome de *Francisella tularensis* subsp. *holarctica* LVS entre 3 et 20 kDa avec ajout de la calibration interne, dont 5 pics sont visibles et annotés en rouge.



**Figure 14** : Spectre MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS acquis sur la gamme de  $m/z$  variant de 3 000 à 20 000, avec calibration interne obtenu avec un MALDI-TOF Biflex IV (Brucker). En abscisse :  $m/z$  ; en ordonnée : intensité en unité arbitraire. En rouge les pics correspondants au calibrant, en noir : les pics du protéome de *Francisella tularensis* subsp. *holarctica* LVS. La ligne de base est soustraite avec le logiciel FlexAnalysis.

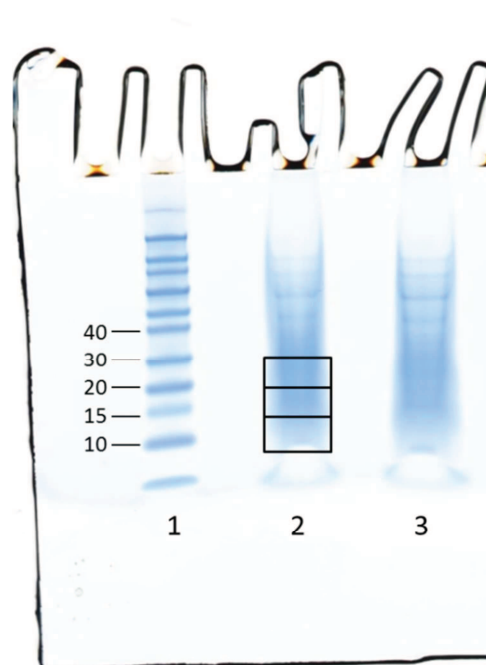
Les données obtenues par spectrométrie de masse MALDI-TOF sont traitées avec le logiciel FlexAnalysis. Ce logiciel permet de réaliser la calibration interne des échantillons, puis d'extraire la liste des rapports masse sur charge ( $m/z$ ) des protéines. Pour notre étude, sont extraits après soustraction de la ligne de base, les 100 rapports  $m/z$  les plus intenses dans la gamme de  $m/z$  de 3 000 à 20 000. La méthode d'extraction des rapports  $m/z$  est la méthode dite « centroïde » qui permet de calculer la masse moyenne à partir des dérivées premières et secondes du signal. Seuls les pics dont le rapport signal sur bruit est supérieur à 6 sont considérés. Pour les  $m/z$  entre 3 000 et 4 000, le bruit de fond peut être très important. Par exemple pour le spectre présenté en **Figure 14**, 67 des 100 rapports  $m/z$  les plus intenses étaient dans la gamme de rapports  $m/z$  inférieurs à 4 000 avant la soustraction de la ligne de base. Après la soustraction de la ligne il n'y en avait plus que 13. En faisant la soustraction de la ligne de base avant l'extraction des rapports  $m/z$ , on obtient un nombre plus important de rapports dans les hautes masses. Les données qui seront traitées par la suite auront subi la soustraction de la ligne de base. Pour déterminer les protéines les plus intenses dans les échantillons inactivés au TFA et soumis au MALDI-TOF, les 100 valeurs de  $m/z$  les plus

intenses ont été listées à partir de quatre analyses (ech311-f2-2, ech311-f2-4, ech313-f4-2, ech313-f4-5). Pour être retenu, un rapport  $m/z$  doit être présent dans au moins 3 listes sur 4.

En classant dans l'ordre croissant les rapports  $m/z$  et leur intensité respective, on obtient 44 valeurs de  $m/z$  détectées dont trois qui correspondent à des pics de calibrant. La liste de ces pics est présentée en **Annexe 1**.

## 2.2 Gel 1D, digestion in-gel et analyse nanoLC-MS/MS

Les échantillons inactivés au TFA sont également résolus sur gel d'électrophorèse. Le marqueur de taille, le gel et le tampon de migration choisis permettent une bonne séparation des protéines de faibles masses. Trois bandes de gel sont découpées de manière à analyser par spectrométrie de masse les protéines de masses inférieures à 30 kDa tel qu'indiqué sur la **Figure 15**. Les protéines contenues dans le gel sont digérées à la trypsine suivant le protocole décrit dans le Matériels et Méthodes (§ b. Protéolyse in-gel).

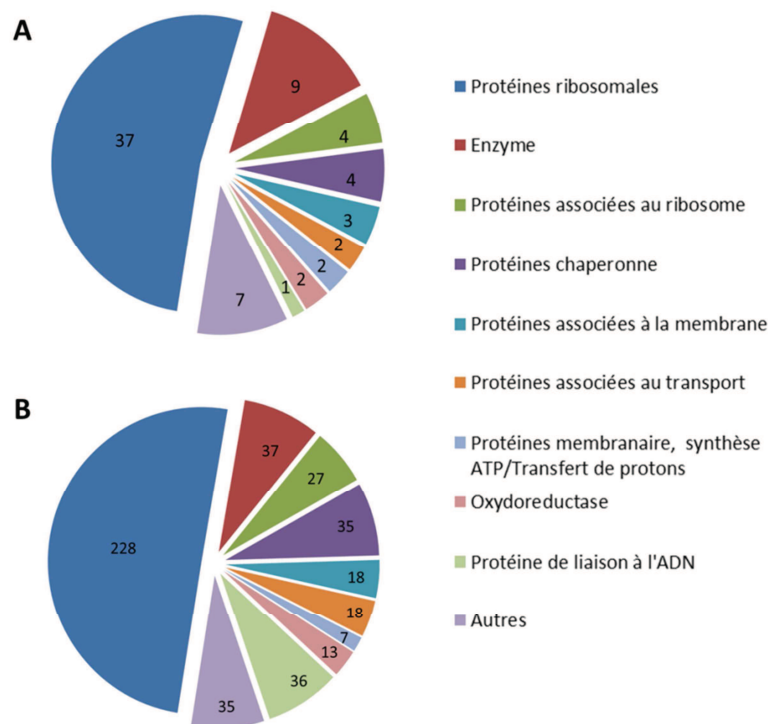


**Figure 15** : Gel d'électrophorèse 1D du protéome de *Francisella tularensis* subsp. *holarctica* LVS inactivée au TFA. Gel NuPAGE®Bis-Tris 4-12% tampon MES, migration à 200 V constant pendant 25 min. Puits 1: 10  $\mu$ l de marqueur de taille Novex Sharp Pre-stained Protein standard. Puits 2 et 3: 15  $\mu$ l de protéome de *Francisella tularensis* subsp. *holarctica* LVS inactivée au TFA.

L'analyse par nanoLC-MS/MS sur un spectromètre de masse LTQ-Orbitrap XL a permis d'acquérir 1 724 spectres MS/MS. Par l'intermédiaire de l'algorithme d'identification Mascot, ces spectres ont été soumis à la base de données nommée FRATH qui contient l'ensemble des séquences protéiques recensées (3 838 au total) dans la base de données NCBI nr pour la bactérie *Francisella tularensis* subsp. *holarctica* LVS à la date du 10 janvier 2012. La base FRATH contient donc des séquences vérifiées et des séquences de protéines prédites. Au total 535 spectres (soit 31% des spectres MS/MS acquis) ont pu être assignés à 282 peptides différents qui valident la présence de 71 protéines avec les paramètres



suivants : peptides tryptiques avec un maximum de deux défauts de coupure peptidique, une tolérance sur la masse de l'ion parent de 5 ppm et de 0,5 Da sur les spectres MS/MS, la carbamidométhylation des cystéines comme modification fixe (+ 57,02 Da) et l'oxydation des méthionines comme modification variable (+ 15,99 Da). L'ensemble des protéines identifiées sont ensuite regroupées selon leur fonction pour une analyse simplifiée. La distribution des protéines selon leur fonction biologique est illustrée dans la **Figure 16** et les références des protéines identifiées dans la base de données sont décrites dans l'**Annexe 2**.



**Figure 16** : Protéines identifiées par nanoLC-MS/MS et caractéristiques fonctionnelles. **A** : Nombre de protéines identifiées selon les groupes fonctionnels. **B** : Nombre de spectres associés aux protéines identifiées selon les groupes fonctionnels.

La part de protéines la plus représentée dans le protéome de *Francisella tularensis* analysé par nanoLC-MS/MS – environ 50 % (228 spectres sur 454) du protéome détecté – correspond aux protéines ribosomales avec 37 protéines sur 71, soit 52 % des protéines identifiées. A part ce groupe fonctionnel, il n'apparaît pas d'autres groupes fonctionnels fortement représentés. La protéine de liaison à l'ADN HU forme B est la seule représentée dans sa catégorie mais est abondamment détectée malgré son faible poids moléculaire (9 468,1 Da) avec 36 spectres (soit 8%). Ceci n'est pas surprenant car cette protéine a plusieurs fois été décrite pour son abondance dans plusieurs bactéries (Christie-Oleza et al., 2013; Toueille et al., 2012). Nous pouvons donc considérer que les résultats obtenus ici peuvent être représentatifs de l'abondance relative de ces protéines dans d'autres bactéries.

### 2.3 Masses des protéines entières et modifications post-traductionnelles

A partir de la liste des séquences d'acides aminés des protéines identifiées, une liste rassemblant l'ensemble des masses des protéines a été créée. La masse isotopique de la protéine est calculée par le logiciel MASCOT à partir de la séquence théorique décrite dans la base de données, alors que le spectre MALDI-TOF va permettre d'établir le rapport  $m/z$  de la protéine, donc sa masse moyenne. De plus, des modifications post-traductionnelles ou de mauvaises annotations sont possibles par rapport à la séquence théorique de la protéine. Ceci peut entraîner une modification de la masse théorique de la protéine. Les masses moyennes des polypeptides ont été calculées à partir des séquences théoriques à l'aide de l'outil « Compute pI/MW » disponible sur la plate-forme ExPASy ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)) et en tenant compte des modifications possibles de ces polypeptides.

La liste constituée contient donc la masse isotopique et les masses moyennes des protéines avec les modifications acétylation (+ 42,0 Da) de l'extrémité N-terminale du polypeptide et/ou perte de méthionine (- 131.2 Da) et pour ces protéines ionisées, leurs rapports  $m/z$  monochargés et dichargés (**Tableau 2**). D'autres modifications post-traductionnelles sont écartées telles que les méthionines oxydées qui sont en proportion faible par rapport aux méthionines classiques, les glycosylations et les phosphorylations peu présentes chez les organismes procaryotes et trop complexes à prévoir.

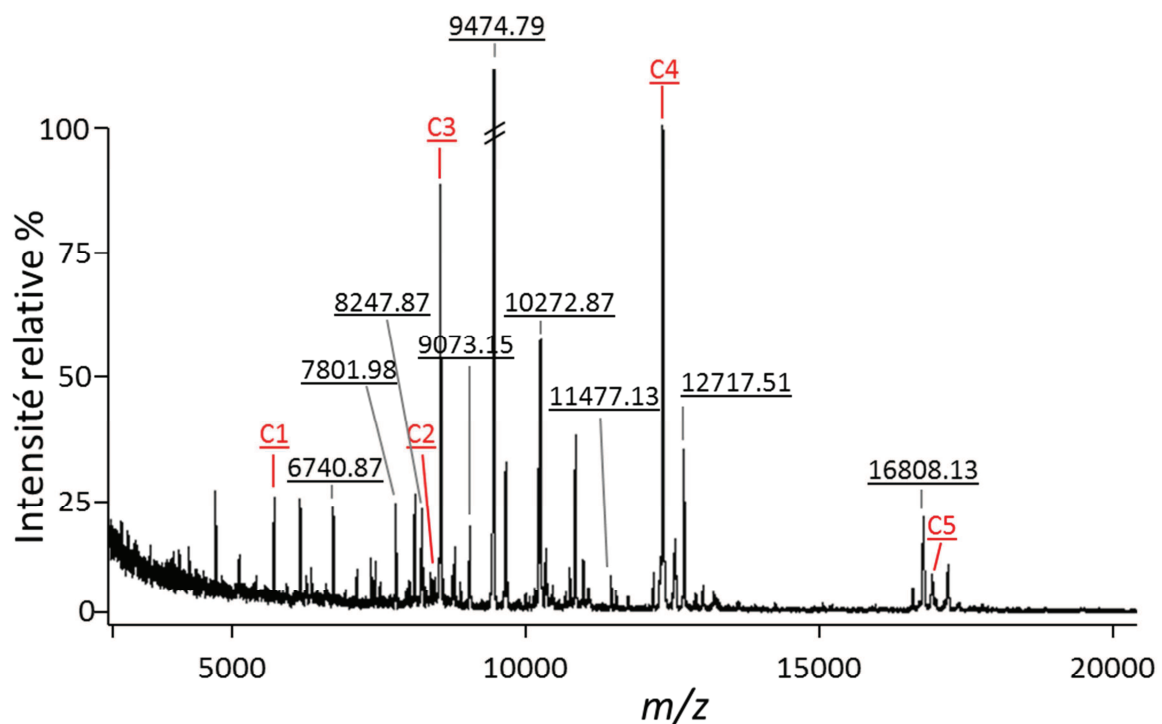
**Tableau 2** : Mode de calcul des masses moyennes des protéines. Acétylation : + Ac. Perte de la méthionine : - Meth.

	Protéine	Masse moyenne
	M	MW
<b>Monochargé</b>	$[M+H]^+$	MW+1
	$[M+H+Ac]^+$	MW+1+42.027
	$[M+H-Meth]^+$	MW+1-131.198
	$[M+H+Ac-Meth]^+$	MW+1+42.027-131.198
<b>Dichargé</b>	$[M+2H]^{2+}$	$(MW+2)/2$
	$[M+2H+Ac]^{2+}$	$(MW+42.027+2)/2$
	$[M+2H-Meth]^{2+}$	$(MW-131.198+2)/2$
	$[M+2H+Ac-Meth]^{2+}$	$(MW+42.027-131.198+2)/2$

### 2.4 Identification des protéines majoritaires

Après comparaison de la liste de masses obtenue par l'analyse MALDI-TOF avec le tableau contenant toutes les masses possibles des protéines identifiées en nanoLC-MS/MS (**Tableau 2**), 18 valeurs de  $m/z$  ont pu être attribuées à des protéines, avec une erreur moyenne de 100 ppm, et une erreur inférieure à 264 ppm dans le pire des cas. Parmi ces 18 valeurs de  $m/z$ , 2 étaient attribuées à deux protéines différentes. La valeur de  $m/z$  6 178 est attribuée avec des erreurs de 214 et 265 ppm. N'ayant aucun élément permettant de choisir l'une des

deux attributions préférentiellement et les deux erreurs étant importantes comparées à l'erreur moyenne sur l'ensemble des pics attribués, nous avons choisi de ne pas retenir ce signal. Pour le second signal de  $m/z$  5137 attribué à deux protéines : la première est identifiée sous sa forme doublement chargée (déjà identifiée sous sa forme monochargée) avec une erreur de 20 ppm, la seconde est identifiée avec une erreur de 184 ppm. Nous avons décidé de retenir logiquement la première attribution. Au final, 15 signaux  $m/z$  attribués à 12 protéines différentes avec une erreur sur la masse inférieure à 105 ppm ont été conservés (**Figure 17**). Trois de ces protéines sont observées à la fois sous leur forme monochargée et doublement chargée, quatre ont subi une perte de la méthionine et deux sont acétylées (**Tableau 3**). L'analyse des séquences d'acides aminés des protéines confirme l'attribution des pertes de méthionines à certaines protéines. En effet, lorsque le second acide aminé est une alanine, une glycine, une proline, une sérine ou une thréonine, il y a toujours perte de la méthionine (Demirev et al., 2001). Pour B2, la thréonine est le deuxième acide aminé, pour B3 et B8 il s'agit de la sérine et B11 présente une alanine en deuxième position. Pour les 8 autres protéines identifiées, aucun des 5 acides aminés cités précédemment n'est en deuxième position dans les séquences des protéines.



**Figure 17** : Spectre MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS avec calibration interne. Les masses des biomarqueurs sont en noir, les pics de calibrant sont repérés en rouge.

**Tableau 3** : Liste des 12 protéines identifiées. \*SC : Spectral Count, nombre de spectres attribués à chaque protéine.

Nom	n° NCBI	Fonction	SC*	Maturation	Charge	<u>m/z théorique</u>		<u>m/z expérimental</u>	Erreur (ppm)
						Isotopique	Moyen	moyen	
B1	CAJ79334.1	Histone-like protein HU form B	36	-	+1	9468.1	9474.9	9475.5	63
B2	CAJ78695.1	50S ribosomal protein L30	21	- Méthionine	+1	6866.7	6740.9	6741.2	39
B3	CAJ79464.1	30S ribosomal protein S18	12	- Méthionine	+1	8373.5	8248.6	8249.0	44
B4	CAJ80177.1	30S ribosomal protein S16	12	-	+1	9066.1	9072.7	9073.5	84
B5	P94797.1	10 kDa chaperonin Cpn10	12	-	+1	10265.4	10272.9	10273.9	105
				-	+2		5136.9	5136.8	-21
B6	CAJ79832.1	Peptidyl-prolyl cis-trans isomerase	10	+ Acétylation	+2	10234.1	5142.3	5141.9	-68
B7	CAJ78688.1	50S ribosomal protein L24	9	-	+1	11469.3	11477.3	11478.0	56
B8	CAJ78962.1	50S ribosomal protein L28	6	- Méthionine	+1	8933.9	8809.2	8810.1	93
B9	CAJ79057.1	Bacterio ferritin	9	-	+1	16798.7	16810.4	16809.9	-27
				-	+2		8405.7	8406.1	46
B10	CAJ78685.1	50S ribosomal protein L29	3	-	+1	7796.4	7802.2	7802.7	66
B11	CAJ80184.1	50S ribosomal protein L7/L12	4	- Méthionine	+1	12839.7	12717.6	12718.5	69
				- Méthionine	+2		6359.3	6359.3	2
B12	CAJ79085.1	Hypothetical lipoprotein	2	+ Acétylation	+2	14873.5	7463.4	7463.6	26

### 3 Identification de biomarqueurs potentiels par approche protéogénomique

L'alignement des 12 protéines identifiées sur la base de données NCBI par l'algorithme PSI-BLAST permet de voir s'il existe des similarités de séquences avec d'autres polypeptides de la base. Les 12 protéines présentent des taux d'identités variant de 90 % à 100 % avec des protéines issues des espèces ou des sous-espèces de *Francisella*. En avril 2013, 24 souches du genre des *Francisella* avaient leur génome complètement ou partiellement séquencé. Leur séquence de nucléotides est disponible dans « Microbial Genomes » de NCBI, ([http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial\\_taxtree.html](http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html)) sous forme complète (une seule séquence) ou sous forme de contigs (plusieurs séquences pouvant se chevaucher). Les souches sont répertoriées dans le **Tableau 4**.

**Tableau 4** : Liste des souches de *Francisella* dont le génome est partiellement ou complètement séquencé

Espèce	Sous-espèce	Souche
tularensis	tularensis	SCHU S4
		FSC198
		WY96-3418
		FSC033
		MA00-2987
		NE061598
	holarctica	LVS
		FTNF002-00
		URTF1
		FSC200
		257
		OSU18
		FSC022
	mediasiatica	FSC147
novicida	U112	
	FTG	
	FTE	
	Fx1	
	GA99-3549	
	GA99-3548	
	3523	
philomiragia	ATCC25015	
	ATCC25017	
TX077308		

Les séquences des 12 protéines sont recherchées par alignement de séquence de protéine (PSI-BLAST) en prenant comme référence la séquence de la protéine de la souche de *Francisella tularensis* subsp. *holarctica* LVS. Pour chaque souche la masse de ces protéines est calculée. A titre d'exemple, le **Tableau 5** présente les résultats de PSI-BLAST pour la protéine de liaison à l'ADN (histone-like protein HU form B, 90 acides aminés) qui apparaît en première position dans la liste des protéines identifiées dans le **Tableau 3** (B1).

**Tableau 5** : Observation des différences de séquences et de masses pour la protéine HU forme B, selon les espèces de *Francisella*.

Espèce/sous espèces	Nombres de souches ayant la même séquence	Modification sur la séquence	Identité (%)	Masse isotopique (Da)	Δ masse (Da)
<i>Francisella tularensis</i>	14	-	100	<b>9468.1</b>	-
<i>Francisella novicida</i>	4	en AA 82 remplacement d'une <b>S</b> par <b>G</b>	99	<b>9438.1</b>	-30.0
<i>Francisella novicida</i>	1	en AA 38 remplacement de <b>N</b> par <b>S</b> en AA 82 remplacement d'une <b>S</b> par <b>G</b>	98	<b>9411.1</b>	-57.0
<i>Francisella novicida</i>	1	en AA 82 remplacement d'une <b>S</b> par <b>G</b> Séquence de 66 AA au lieu de 90 AA	98	<b>6921.9</b>	-2546.2
<i>Francisella philomiragia</i> et <i>Francisella</i> sp. TX077308	3	en AA 57 remplacement de <b>P</b> par <b>A</b> en AA 72 remplacement de <b>P</b> par <b>A</b> en AA 82 remplacement d'une <b>S</b> par <b>G</b>	97	<b>9386.1</b>	-82.0

En observant précisément les séquences de ces protéines nous pouvons constater que malgré la forte identité des séquences, la modification d'un seul acide aminé va avoir pour conséquence de modifier sensiblement la masse de la protéine. Dans le cadre de notre exemple, un changement de masse de 30 Da entre la séquence de la protéine d'une souche de *Francisella tularensis* et d'une souche *Francisella novicida* sera observable sur un spectre MALDI-TOF.

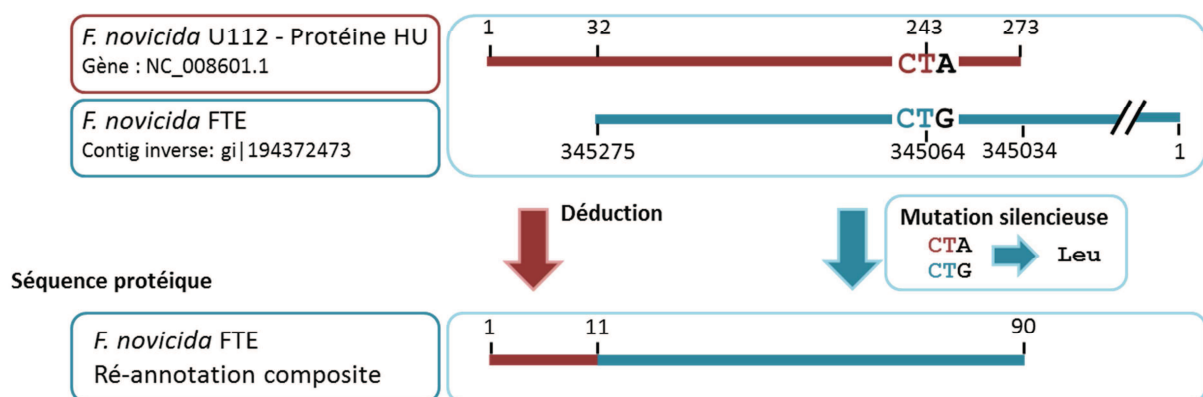
Dans le **Tableau 5**, il apparaît également deux anomalies. Premièrement, malgré le fait que cette séquence d'acides aminés semble très conservée au sein du genre des *Francisella*, la souche de *Francisella novicida* GA99-3548 présente pour cette protéine une séquence de 66 acides aminés au lieu des 90 acides aminés attendus. Deuxièmement, lors de l'alignement de la séquence de la protéine contre la base NCBI, aucune séquence n'a été trouvée pour la souche *Francisella novicida* FTE. Il se pose donc la question suivante : cette protéine est-elle absente du protéome de *Francisella novicida* FTE ? Pour comprendre, une étude protéogénomique des 12 protéines pour les 24 souches a été entreprise pour réannoter le génome par les données protéomiques.

La démarche est la suivante : pour chaque protéine manquante lors de l'alignement de séquence ou dont la masse est sensiblement différente de celle des autres protéines de la même espèce ou sous-espèce, une recherche de la séquence du gène va être menée. Pour cela il faut choisir des organismes de références. Pour la sous-espèce *tularensis holarctica*, il

s'agit de la souche LVS. Pour la sous-espèce *tularensis tularensis*, il s'agit de la souche SCHU S4. Pour l'espèce *novicida*, la référence est U112. Les autres espèces et sous-espèces étant peu peuplées, elles sont confrontées à la souche de référence de cette étude, la souche LVS.

La **Figure 18** montre sur la souche FTE la démarche protéogénomique réalisée pour proposer une nouvelle annotation de la protéine HU forme B (B1 dans le **Tableau 3**). Le gène de la protéine HU forme B de *F. novicida* U112 est aligné sur le génome de *F. novicida* FTE (sous forme de contigs) par un alignement de nucléotides (BLASTn). Alors que son protéome ne possédait aucune protéine correspondant à la protéine HU forme B, 242 nucléotides sont reconnus à l'extrémité du contig gi| 194372473. Une mutation apparaît dans la séquence (CTG au lieu de CTA). Cette mutation est silencieuse au moment de la traduction car ces deux codons codent pour la leucine. La séquence étant en extrémité de contig et le taux d'identité des acides aminés étant de 100 % sur cette séquence, nous proposons d'ajouter la protéine HU forme B au protéome de *F. novicida* FTE avec comme séquence, celle complète de la protéine HU forme B de la souche U112.

#### Séquence de nucléotide



**Figure 18** : Démarche protéogénomique. La séquence de nucléotide du gène recherché est alignée sur le génome de la souche. Une partie du gène est reconnue sur une extrémité de contig. Une proposition de ré-annotation de la protéine est réalisée sur la base de la séquence de la protéine de référence.

Au total, sur les 288 masses de protéines recherchées, 24 n'ont pas pu être trouvées par requête PSI-BLAST. Les recherches sur les génomes ont permis d'identifier 4 protéines dont la séquence de nucléotide partielle ou complète était non annotée. La protéine HU forme B (B1 dans le **Tableau 3**) de *Francisella novicida* GA99-3548, dont la séquence nucléotidique était annotée, était présente en bout de contig, résultant en une séquence de protéine tronquée. Nous avons proposé de la ré-annoter avec la séquence de référence correspondante de *Francisella novicida* U112. Ensuite, les protéines ribosomales 30S 16S (B4 dans le **Tableau 3**) de *F. t.* subsp. *tularensis* FSC033 et de *F. novicida* Fx1 comportaient des démarrages en N-terminal différents des autres bactéries de la même sous-espèce. Dans ces cas, la séquence de nucléotide est identique et présente en entier sur le génome cible mais la protéine annotée est plus courte que la protéine de référence. Nous avons donc proposé

de ré-annoter ces deux protéines avec comme séquence, la séquence de référence (*F. t. subsp. tularensis* SCHU S4 et *F. novicida* U112 respectivement).

Plusieurs groupes de protéines révèlent avoir des masses identiques alors qu'il existe des différences dans leur séquence d'acides aminés.

Premièrement, les souches ATCC 25015 et ATCC 25017 ont leur peptidyl-prolyl cis-trans isomerase (B6 dans le **Tableau 3**) de même masse alors qu'il existe une différence dans leur séquence d'acides aminés. En effet, la souche ATCC 25017 voit une glutamine remplacée par une lysine, la différence de masse n'est que de 0,04 Da, différence difficile à détecter en spectrométrie de masse MALDI-TOF.

Deuxièmement, les protéines ribosomales 30S 16S (B4 dans le **Tableau 3**) de *F. novicida* GA99-3548 et des deux souches de *F. philomiragia* (séquences identiques pour les deux souches) présentent les mêmes masses. Entre les deux séquences, en position 52 et 81 une alanine remplace une valine et inversement une valine remplace une alanine. Il n'y a donc pas de changement de masse visible entre les protéines, en revanche, lors d'une étude des peptides ces modifications seraient discriminantes.

Le dernier groupe de 9 protéines ayant la même masse concerne la protéine ribosomale 50S L7/L12 (B11 dans le **Tableau 3**) de 6 souches de la sous-espèce *holarctica* (toutes les souches sauf FSC022) et trois souches de la sous-espèce *tularensis* (SCHU S4, FSC198 et NE061598). Dans chaque sous-espèce, la séquence de la protéine est identique mais entre les deux sous-espèces apparaissent deux modifications : en position 50 et 110 une alanine remplace une valine et inversement une valine remplace une alanine. De nouveau les masses sont identiques malgré des modifications dans la séquence des polypeptides.

Au final, 9 masses de protéines ne peuvent être déterminées dans le **Tableau 6**, certainement parce que les génomes ne sont pas complètement séquencés. En effet, l'une des premières étapes de la constitution de la séquence complète du génome est de créer des contigs à partir des fragments de génomes séquencés. Les algorithmes se basent sur le recouvrement des fragments pour créer les contigs, il est donc possible que certains fragments ne soient pas encore intégrés au contigs disponibles ou n'aient pas été amplifiés avant l'étape de séquençage. Quatre autres protéines ne sont pas trouvées alors que le génome est entièrement séquencé. Soit ces protéines ne sont pas exprimées dans ces organismes, soit il s'agit d'erreurs au moment du séquençage.

L'ensemble des masses obtenues est rassemblé dans le **Tableau 6**.



**Tableau 6** : Liste des masses de 12 biomarqueurs pour les 24 souches de *Francisella*

Espèce	Sous-espèce	Souche	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
<i>tularensis</i>	<i>holarctica</i>	LVS	9473.90	6871.10	8378.82	9071.71	10271.87	10240.55	11476.34	8939.44	16809.36	7801.20	12847.82	14882.72
		FTNF002-00	9473.90	6871.10	8378.82	9071.71	10271.87	10240.55	11476.34	8939.44	18463.39	7801.20	12847.82	14882.72
		URTF1	9473.90	6871.10	8378.82	9071.71	10271.87	10240.55	11476.34	8939.44	17992.72	7801.20	12847.82	14882.72
		FSC200	9473.90	6871.10	8378.82	9071.71	10271.87	10240.55	11476.34	8939.44	16809.36	7801.20	12847.82	14882.72
		257	9473.90	6871.10	8378.82	9071.71	10271.87	10240.55	11476.34	8939.44	18463.39	7801.20	12847.82	14882.72
		OSU18	9473.90	6871.10	8364.79	9071.71	10271.87	10240.55	11476.34	8939.44	18463.39	7801.20	12847.82	14882.72
	<i>tularensis</i>	FSC022	9473.90	6901.13	8411.83	9071.71	10271.87	10240.55	11476.34	8939.44	18507.35	7773.15	*	14896.75
		SCHU S4	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	11534.38	8939.44	18507.35	7773.15	12847.82	14882.72
		FSC198	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	11534.38	8939.44	18507.35	7773.15	12847.82	14882.72
		WY96-3418	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	11534.38	8939.44	18036.68	7773.15	12819.79	14882.72
		FSC033	9473.90	6871.10	8351.79	9081.74 <sup>3</sup>	10244.85	10240.55	11534.38	8939.44	18507.35	7773.15	*	14882.72
		MA00-2987	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	*	8939.44	18507.35	*	*	14882.72
	<i>mediasiatica</i>	NE061598	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	11534.38	8939.44	18036.68	7773.15	12847.82 <sup>6</sup>	14882.72
		FSC147	9473.90	6871.10	8351.79	9081.74	10244.85	10240.55	11476.34	8939.44	16795.29	7773.15	12920.87	14912.75
<i>novicida</i>	U112	9443.87	6871.10	8351.79	9081.74	10244.85	10240.55	11476.34	8939.44	16795.29	7773.15	12819.79	14853.64	
	FTG	9443.87	6871.10	8351.79 <sup>1</sup>	9081.74	10244.85	10240.55	11476.34	8939.44	18036.68	*	12819.79	14853.64	
	FTE	9443.87 <sup>1</sup>	6871.10	8351.79 <sup>1</sup>	9081.74	10244.85	10240.55	11476.34	8939.44	18036.68	7773.15 <sup>1</sup>	12819.79	14853.64	
	Fx1	9443.87	6871.10	8351.79	9081.74 <sup>3</sup>	10244.85	10240.55	11476.34	8939.44	16795.29	7815.23	12819.79	14881.69	
	GA99-3549	9443.87	6871.10	8351.79	9081.74	10244.85	10217.52	11476.34	8939.44	18507.35	7773.15	*	14853.64	
	GA99-3548	9443.87 <sup>2</sup>	6871.10	8351.79	9109.80 <sup>5</sup>	10244.85	10217.52	11476.34	8939.44	18507.35	7773.15	12819.79	14839.61	
	3523	9416.84	6871.10	8351.79	8510.99	10175.70	10213.49	11492.34	8997.48	16922.57	7773.15	12819.79	14924.67	
<i>philomiragia</i>	ATCC25015	9391.79	6873.07	8337.76	9109.80	10247.72	10170.50 <sup>4</sup>	11448.29	8941.45	*	7757.15	12738.65	*	
	ATCC25017	9391.79	6873.07	8337.76	9109.80	10247.72	10170.54 <sup>4</sup>	11448.29	8941.45	#	7757.15	12738.65	#	
<i>TX077308</i>	Sp.TX077308	9391.79	6873.07	8337.76	9081.74	10248.72	10156.46	11434.28	8895.42	#	7773.15	12795.70	#	
Number of distinct polypeptides			4	3	5	5	5	6	5	4	7	4	5	6

\* Séquence codante non annotée, probablement dû au génome non complet (séquençage non terminé).

# Séquence codante non annotée alors que le génome est complet.

<sup>1</sup> Protéine non annotée dans NCBIInr jusqu'à présent. Une séquence codante partielle a été trouvée en bout de contig du génome par alignement de séquence de nucléotide avec le gène d'un homologue, mais il n'y a pas de protéine annotée correspondant à cette séquence codante. Pas de mutations détectées au niveau de la séquence de nucléotide qui provoque un changement dans la séquence d'acides aminés de la protéine (ou mutation silencieuse). Compte tenu de cela, il est raisonnable de penser que ces protéines existent probablement dans le génome de la bactérie avec la même séquence protéique et donc la même masse, que son plus proche homologue.

<sup>2</sup> Protéine mal annotée dans NCBIInr. La protéine annotée jusqu'à présent, est en fait seulement partiellement encodé sur un gène situé à l'extrémité d'un contig du génome comme le montre l'alignement de séquence d'un gène homologue avec le génome de la souche. Pas de mutations détectées au niveau de la séquence de nucléotide qui provoque un changement dans la séquence d'acides aminés de la protéine (ou mutation silencieuse). Compte tenu de cela, il est raisonnable de penser que ces protéines existent probablement dans le génome de la bactérie avec la même séquence protéique et donc la même masse, que son plus proche homologue.

<sup>3</sup> Protéine mal annotée dans NCBIInr au niveau du N-terminal. La comparaison de la séquence de la protéine avec celles des autres bactéries de la même sous-espèce montre une conservation de la séquence sur la totalité de la longueur mais les premiers acides aminés n'apparaissent pas comme dans la séquence de référence. La présence d'un codon initiateur alternatif a été documentée, et il existe dans le génome une séquence codante identique au gène des autres bactéries de la même sous espèce, sur toute la longueur. Compte tenu de cela, il est raisonnable de penser que ces protéines existent probablement dans le génome de la bactérie avec la même séquence protéique et donc la même masse, que son plus proche homologue.

<sup>4</sup> Protéines montrant des différences en termes de séquence d'acides aminés (*F. philomiragia* souches ATCC25015 et ATCC25017), mais dont les masses résultantes sont identiques.

<sup>5</sup> Protéines montrant des différences en termes de séquence d'acides aminés en comparaison avec les souches de *F. philomiragia*, mais dont les masses résultantes sont identiques.

<sup>6</sup> Protéines montrant des différences en termes de séquence d'acides aminés en comparaison avec les autres souches de *F. tularensis holarctica* et *F. tularensis* SCHU4 et FSC198, mais dont les masses résultantes sont identiques.

#### 4 Choix des biomarqueurs

Le genre *Francisella* possède plusieurs sous-espèces de pathogénicités différentes pour l'homme. L'espèce la plus pathogène pour l'homme est *Francisella tularensis*. Il est donc important de pouvoir déterminer si une *Francisella* est présente dans un échantillon et le cas échéant de pouvoir déterminer l'espèce et la sous-espèce auxquelles la bactérie appartient. Comme le montre le **Tableau 6** certaines protéines sont très conservées au sein des différentes espèces de *Francisella*, comme la protéine ribosomale 50S L30 (B2). En effet pour cette protéine seulement deux séquences différentes existent parmi les 24 souches de *Francisella* séquencées. D'autres protéines au contraire, comme la ferritine (B9) et une lipoprotéine (B12), ont des séquences très différentes selon les souches avec 6 ou 7 séquences protéiques différentes, sans compter les souches pour lesquelles elles n'ont pas de séquences codantes. L'observation de cette table montre également qu'aucune protéine ne peut permettre à elle seule de distinguer toutes les souches. Nous avons donc mis au point une stratégie d'identification des différentes espèces et sous-espèces de *Francisella* reposant sur la reconnaissance des masses de trois protéines. En effet, comme aucune protéine seule ne donnait la possibilité de réaliser ce type d'identification, nous avons observé protéine par protéine lesquelles étaient conservées pour toutes les souches dans une même sous-espèce ou espèce. La protéine B1 permettait dans un premier temps de différencier les espèces *tularensis*, *novicida* et *philomiragia*. Comme les séquences polypeptidiques sont très conservées au sein de l'espèce *tularensis*, ce sont deux nouvelles protéines qu'il a fallu choisir pour réussir à distinguer les trois sous-espèces. Au final, les protéines choisies furent : la protéine HU forme B (B1), la protéine chaperonne Cpn10 (B5) et la protéine ribosomale 50S L24 (B7). La combinaison des masses de ces trois protéines permet de différencier à la fois les espèces de *Francisella* et les sous espèces de *Francisella tularensis* comme le montre le **Tableau 7**.

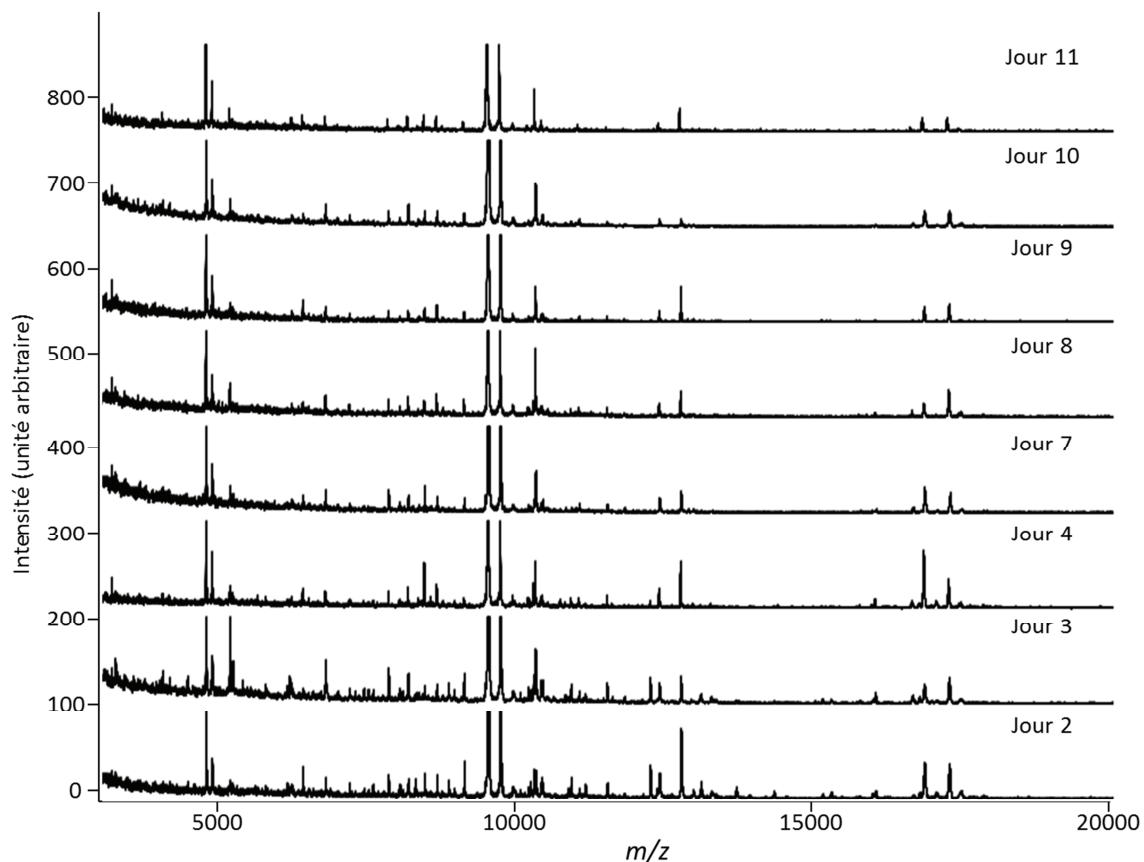
**Tableau 7** : Combinaisons de masses spécifiques aux espèces et aux sous-espèces de *Francisella* (masses moyennes monochargées, telles qu'attendues sur un spectre MALDI-TOF)

Espèce	Sous-espèce	B1	B5	B7
<i>F. tularensis</i>	<i>holarctica</i>	9474.9	10272.9	11477.3
	<i>tularensis</i>	9474.9	10245.5	11535.4
	<i>mediasiatica</i>	9474.9	10245.5	11477.3
<i>F. novicida</i>		9444.9	10245.5	11477.3
		9417.8	10176.7	11493.3
<i>F. philomiragia</i>		9392.8	10248.7	11449.3
<i>F. TX077308</i>		9392.8	10249.7	11435.3

## 5 Validation des biomarqueurs

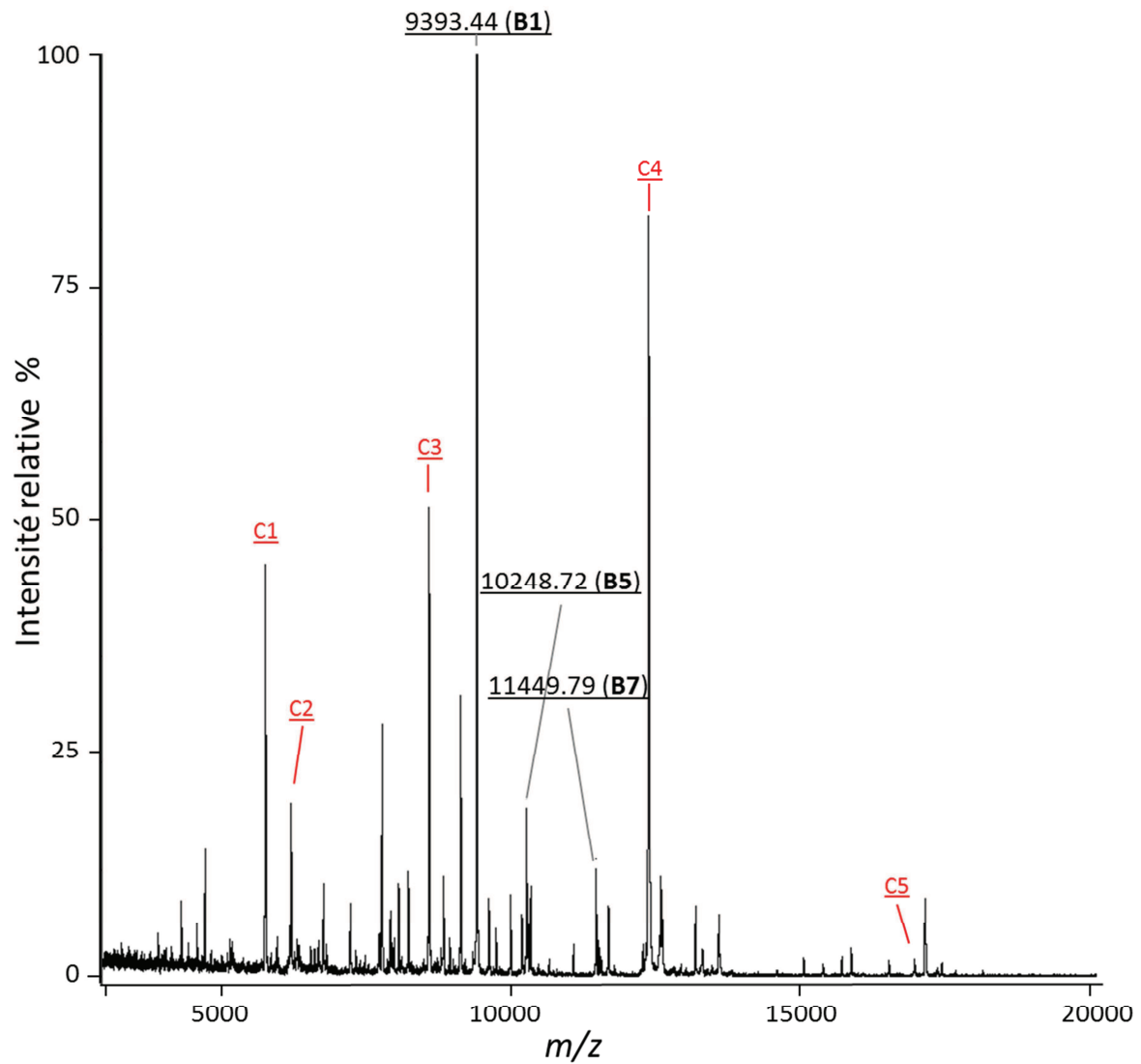
Pour valider le choix de nos trois biomarqueurs nous nous sommes basés sur la littérature et sur des validations pratiques avec différentes souches de *Francisella*.

Tout d'abord nous avons vérifié la stabilité du protéome détecté par MALDI-TOF au cours du temps. Pour cela, nous avons mis en culture *Francisella tularensis* subsp. *holarctica* LVS sur une boîte de gélose chocolat + (PVX) pendant plusieurs jours à 37°C. Aux jours 2, 3, 4, 7, 8, 9, 10 et 11 nous avons fait des prélèvements de colonies suivis d'une inactivation au TFA. Ces échantillons inactivés ont été analysés par MALDI-TOF et les profils comparés (**Figure 19**). Aucune différence majeure n'est observée entre ces 8 profils hormis une diminution de l'intensité des pics au cours du temps.

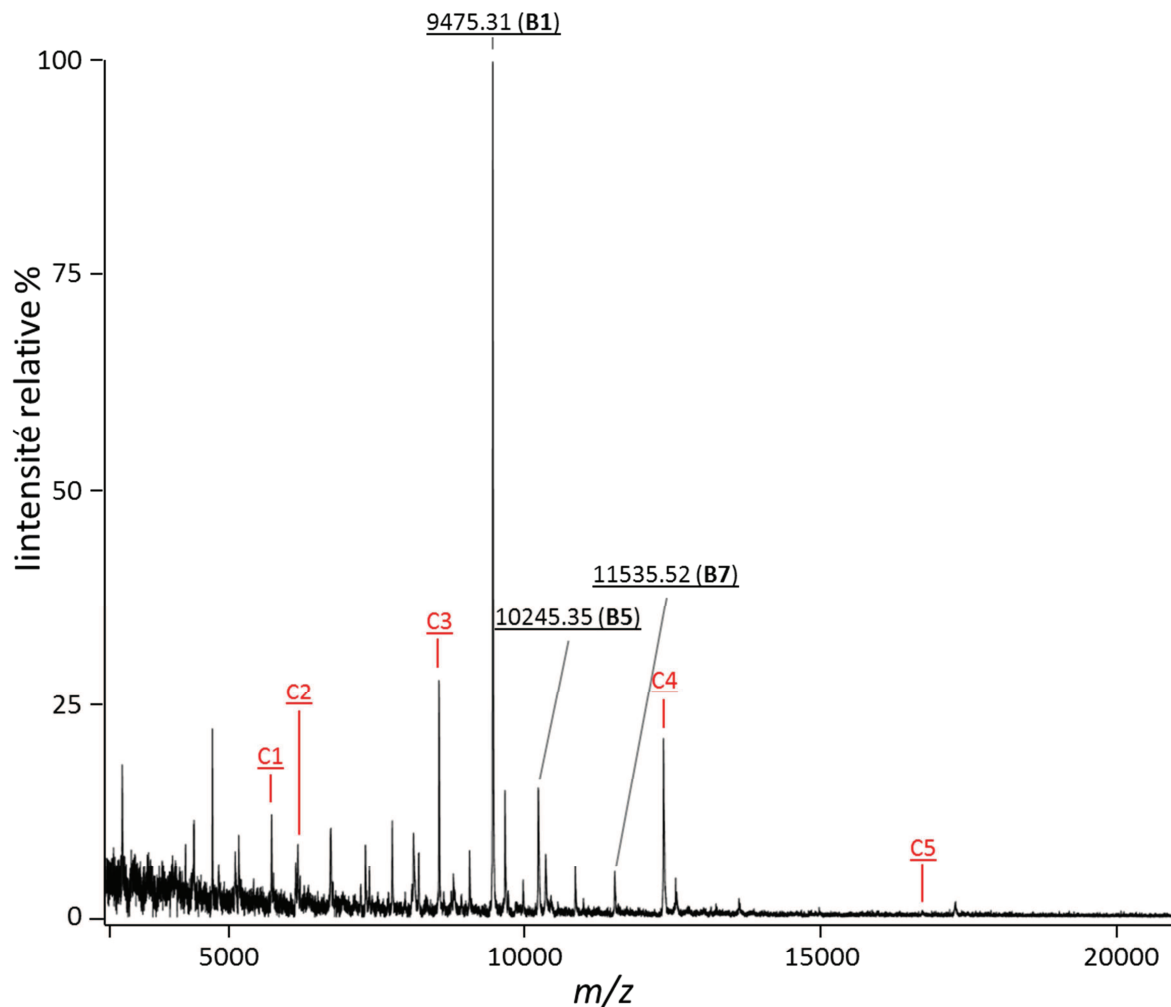


**Figure 19** : Evolution des profils MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS au cours de 11 jours de culture à 37°C.

Nous avons ensuite réalisé les spectres MALDI-TOF de deux souches de *Francisella* pour valider la prédiction des masses des biomarqueurs. Les spectres de *Francisella philomiragia* ATCC 25015 et *Francisella tularensis* subsp. *tularensis* CCUG 2112 sont représentés en **Figure 20** et en **Figure 21** respectivement.



**Figure 20** : Spectre MALDI-TOF de *Francisella philomiragia* ATCC 25015. En rouge sont représentés les pics du calibrant interne, en noir, les masses des biomarqueurs B1, B5 et B7.



**Figure 21 :** Spectre MALDI-TOF de *Francisella tularensis* subsp. *tularensis* CCUG 2112. En rouge, sont représentés les pics du calibrant interne, en noir, les masses des biomarqueurs B1, B5 et B7.

Comme attendu nous retrouvons bien les masses des protéines B1, B5 et B7 sous leur forme monochargée. Avec la calibration interne nous obtenons les rapports  $m/z$  suivants pour *F. philomiragia* ATCC 25015 : 9 393,4, 10 249,0 et 11 449,8 avec une erreur de moins de 63 ppm sur les masses prédites (**Figure 20**). Pour *F. tularensis* subsp. *tularensis* CCUG 2112, nous obtenons les rapports  $m/z$  9 475,3, 10 245,4 and 11 535,5 avec une erreur de moins de 42 ppm sur les masses prédites (**Figure 21**). Ceci confirme bien nos prédictions protéogénomiques en termes de masses. De plus le profil d'intensité des spectres est également semblable quel que soit la souche étudiée. En effet, pour nos trois souches, la protéine majoritaire en termes d'intensité est la protéine HU forme B (B1) et les deux autres protéines cibles contribuent également au motif des signaux de rapport  $m/z$  les plus intenses. Ces résultats montrent que le choix de ces trois biomarqueurs est pertinent puisque les masses prédites sont bien retrouvées chez toutes les souches de *Francisella* étudiées. Une publication de 2010 de l'équipe de Seibold présente l'analyse par « whole cell MALDI-TOF » d'une série de 50 souches de *Francisella* (Seibold et al., 2010). Leur objectif était de proposer des groupes de souches en fonction de leurs caractéristiques spectrales

communes sans pour autant identifier les protéines. Cette équipe a également observé que certaines masses étaient conservées pour toutes les souches, un signal de rapport  $m/z$  5 180 notamment, alors que pour d'autres signaux le rapport  $m/z$  d'une protéine variait de quelques unités selon les espèces ou les sous-espèces. Parmi les spectres proposés dans l'article, nous avons cherché à confirmer la présence des trois biomarqueurs que nous avons définis. Nous avons clairement pu observer les signaux provenant de la protéine HU forme B (B1) sur les spectres de *F. philomiragia* F050, *F. novicida* F061, *F. tularensis mediasiatica* F063, *F. tularensis tularensis* FL3/1 et *F. tularensis holarctica* LVS. Le signal de rapport  $m/z$  aux environs de 9 500 correspond aux masses prédites pour les différentes souches et est également le plus intense. Ces observations confirment une nouvelle fois la validité et la pertinence de notre approche. Malheureusement, les spectres rapportés dans cette publication sont acquis pour des valeurs de  $m/z$  entre 4 500 et 11 000 et les signaux de  $m/z$  au-delà de 10 000 ne sont pas annotés, ce qui ne permet pas de comparer les masses des deux autres biomarqueurs.

## 6 Conclusion

Dans cette partie nous avons proposé un ensemble de trois protéines dont les masses permettent de classer rapidement les souches de *Francisella* en fonction de leur espèce ou de leur sous-espèce le cas échéant. L'étude combinée du protéome de rapports  $m/z$  inférieurs à 20 kDa à la fois au MALDI-TOF et par nanoLC-MS/MS a permis l'identification de protéines entières observées par MALDI-TOF. L'avantage d'identifier ces protéines est de pouvoir prédire les masses des protéines pour les souches dont le génome a été séquencé mais qui ne sont pas forcément à disposition du laboratoire ou dont la manipulation impose des contraintes. L'étude systématique des masses des protéines en fonction de leur séquence d'acides aminés a permis de mettre à jour des erreurs d'annotations dues soit à des génomes incomplets, soit à des erreurs d'annotation du génome. L'ensemble des masses obtenues a montré qu'il était possible d'identifier les espèces et certaines sous-espèces de *Francisella* avec l'observation des masses de trois protéines sur les spectres MALDI-TOF : la protéine HU forme B (B1), la protéine chaperonne Cpn10 (B5) et la protéine ribosomale 50S L24 (B7). Notre étude s'est effectuée sur la souche LVS car elle est manipulable en laboratoire de biosécurité de niveau 2, contrairement aux souches plus virulentes. Comme l'objectif est de différencier les souches des sous-espèces les plus virulentes de celles non pathogènes, il fallait pouvoir vérifier que les masses prédites étaient correctes pour d'autres souches que la LVS et notamment sur la sous-espèce la plus virulente. Au laboratoire était disponible la souche ATCC 25015 de *F. philomiragia*, qui a confirmé une première fois les masses prédites, puis nous nous sommes procuré, par l'intermédiaire de l'université de Göteborg, une souche de *F. tularensis tularensis*, la sous-espèce la plus virulente des *Francisella*. Une nouvelle fois, les masses prédites ont été confirmées. Il est important de noter que les protéines choisies se trouvent dans des gammes de  $m/z$  (9 370-9 480, 10 170-10 280, et 11 430-11 540) où peu d'autres pics sont présents dans les conditions où nous avons réalisé l'analyse. De manière à supprimer les

effets de bruit de fond qui sont nombreux à faibles rapports  $m/z$ , il pourrait être intéressant d'analyser uniquement la gamme de rapports  $m/z$  de 9 000 à 12 000 ou encore trois gammes de rapports  $m/z$  réduites autour des rapports  $m/z$  des biomarqueurs.

Cette approche nécessite d'avoir une culture pure de bactéries car un mélange entraînerait un nombre de pics très important sur les spectres MS, une atténuation ou une suppression de certains pics et une augmentation du bruit de fond. Toutefois, un mélange restreint pourrait être analysé, mais du fait de la grande variété des espèces bactériennes présentes sur terre, il est possible qu'une souche autre qu'une *Francisella* possède ces 3 rapports  $m/z$ . Une étape préliminaire d'isolation sur différents milieux nutritifs ou de confirmation par un test de gram pourrait être utile pour confirmer qu'il s'agit bien d'une *Francisella*, ou des étapes d'enrichissement avec des anticorps spécifiques aux bactéries ou aux protéines cibles de manière à diminuer le nombre de protéines à étudier. De nouvelles méthodes analytiques, comme la nanoLC-MS de protéines entières (mode dit « top-down »), pourraient aussi permettre de chercher ces trois biomarqueurs avec plus de précision et une corrélation supplémentaire provenant du temps de rétention sur la colonne en phase inverse de chaque biomarqueur. La puissance de discrimination du couplage chromatographie et spectrométrie de masse en tandem, ainsi que la détermination de la masse avec une plus grande précision que la technologie MALDI-TOF permettrait l'identification des 3 biomarqueurs même dans le cas d'un mélange complexe. Pour terminer, l'étude combinée en MALDI-TOF-MS et en nanoLC-MS/MS d'un protéome et l'approche protéogénomique développée ici pourrait être étendue à d'autres genres afin d'identifier de nouveaux biomarqueurs.





# 2<sup>ème</sup> partie : Immunocapture de *Francisella tularensis* couplée à une détection MALDI-TOF

## 1 Contexte et stratégie proposée

L'identification de microorganismes par spectrométrie de masse se heurte à deux inconvénients majeurs. Le premier est la sensibilité des appareils de spectrométrie de masse et le second est la pureté de l'échantillon. En effet si l'organisme à détecter et à identifier n'est pas assez concentré, il sera en dessous de la limite de détection de l'appareil. De plus, si le microorganisme est au sein d'un mélange, il peut être difficile de le différencier des contaminants.

L'objectif de cette seconde partie est donc de mettre au point une méthodologie de concentration et de purification d'un microorganisme en milieu complexe (mélange bactérien ou matrice environnementale) afin de faciliter la détection et l'identification par spectrométrie de masse MALDI-TOF. Pour cela nous avons choisi d'utiliser des billes magnétiques couplées à des anticorps ciblant la surface des bactéries entières. Le modèle bactérien choisi pour cette étude est *Francisella tularensis* subsp. *holarctica* LVS, la souche vaccinale de *Francisella tularensis* subsp. *tularensis*, la sous-espèce la plus virulente des *Francisella*.

Les différentes étapes d'optimisation du protocole d'immunocapture magnétique sont détaillées dans cette partie ainsi que la validation du protocole sur plusieurs types de matrices. Il y est également décrit comment sont validés les résultats obtenus par spectrométrie de masse.

## 2 Méthodologie choisie

Comme indiqué dans l'introduction bibliographique, un important travail a déjà été effectué sur l'enrichissement de bactéries avant l'analyse par spectrométrie de masse. De toutes les méthodes étudiées, la méthode à la fois la plus spécifique et la plus sensible est l'immunocapture par billes magnétiques. De plus, au vue de la spécificité et du taux de greffage qu'il est possible d'obtenir avec cette technologie, il semble que ce soit la meilleure solution pour un bon enrichissement. Les billes magnétiques sont alors soit déposées directement sur les cibles MALDI, soit une élution est réalisée et le surnageant est déposé. Avec cette méthode, un grand nombre de matrices et de mélanges complexes ont été étudiés lors de ces recherches, que ce soit des matrices humaines (prélèvements nasaux, buccaux ou urine) ou environnementales (eaux de rivière, prélèvements d'air, etc.).

A propos de *Francisella*, un travail sur l'enrichissement de peptides par immunocapture magnétique a été conduit (Jiang et al., 2007). Un anticorps visant un peptide tryptique de la protéine IgC de l'îlot de pathogénicité de *F. tularensis* (acides aminés 49 à 61) (Santic et al., 2005) est immobilisé sur des billes de sépharose qui sont ensuite incubées durant 2 à 4 heures dans des digestions de plasma humain enrichis avec des peptides de *F. tularensis* ou des digestions bactériennes. Après rinçage, 0,5 µl de la solution de billes resuspendues dans du PBS est déposée sur cible MALDI. Dans le plasma la sensibilité obtenue est de 1 280 CFU/ml. Même si cette méthode est moins sensible dans du plasma que dans du PBS, elle est tout de même comparable aux limites atteintes par les tests ELISA (10<sup>3</sup> CFU/ml). En revanche, aucune étude ne s'est portée sur l'enrichissement à l'aide d'anticorps de *Francisella* dirigés contre la bactérie entière. Nous avons donc entrepris de développer cette approche.

Tout d'abord, il faut choisir la technologie de billes la plus adaptée à notre étude. Le protocole sera ensuite mis au point dans une matrice simple (PBS) et sera validé dans des matrices complexes (mélange de bactéries, eau minérale, lait, prélèvement d'air) avec différentes concentrations de bactéries. Nous devons également établir le seuil de détection de *Francisella* par spectrométrie de masse MALDI-TOF.

### 3 Mise en œuvre des anticorps

#### 3.1 Fabrication des anticorps monoclonaux

Les anticorps monoclonaux ont été conçus et produits par l'équipe de Laurent Bellanger (Laboratoire d'Ingénierie Cellulaire et Biotechnologie, LICB) du SBTN. La souche *Francisella tularensis* subsp. *holarctica* LVS a la particularité de posséder la même enveloppe externe (lipopolysaccharide) que les souches de la sous-espèce la plus virulente, *Francisella tularensis* subsp. *tularensis*. Une solution de *Francisella tularensis* subsp. *holarctica* LVS inactivées à la chaleur a été utilisée pour immuniser par injection intrapéritonéale des souris et des rats en présence d'un adjuvant. L'intérêt de l'inactivation à la chaleur est qu'il n'y a pas d'ajout de produits chimiques donc, i) moins de risques toxiques pour les animaux immunisés, ii) pas de modifications chimiques des antigènes et iii) pas de modifications du lipopolysaccharide (LPS) qui est le composant non protéique de la membrane externe de la bactérie. Avant l'immunisation il a été vérifié que les bactéries étaient correctement inactivées en ensemencant celles-ci sur des boîtes de gélose chocolat. L'inactivation a été validée lorsqu'au bout d'une semaine de culture à 37°C aucune colonie n'est apparue sur la boîte. L'adjuvant de Freund, utilisé pour la préparation de la solution d'immunisation, est un mélange lipidique immunogène qui permet de stimuler l'immunité. Chaque mois, les animaux ont reçu une nouvelle injection de cet immunogène et deux semaines après chaque injection, le sérum de chaque animal immunisé a été prélevé et titré. Les animaux dont la titration était la plus élevée furent sacrifiés et les lymphocytes B prélevés à partir de la rate. Les lymphocytes B ont ensuite été fusionnés avec des myélomes. Ces hybridomes ont été

mis en culture 10 jours à 37°C sous atmosphère humide et à teneur en CO<sub>2</sub> contrôlée. Les surnageants de cultures ont été testés par tests ELISA sur des *Francisella tularensis* subsp. *holarctica* LVS inactivées puis sur des *Francisella tularensis* subsp. *holarctica* LVS vivantes avant clonage pour rechercher la présence d'anticorps monoclonaux. Les anticorps choisis ont également été testés contre d'autres bactéries telles que *E. coli* et *Francisella philomiragia*, une autre espèce de *Francisella* pour vérifier qu'ils ne présentaient pas de réactions croisées.

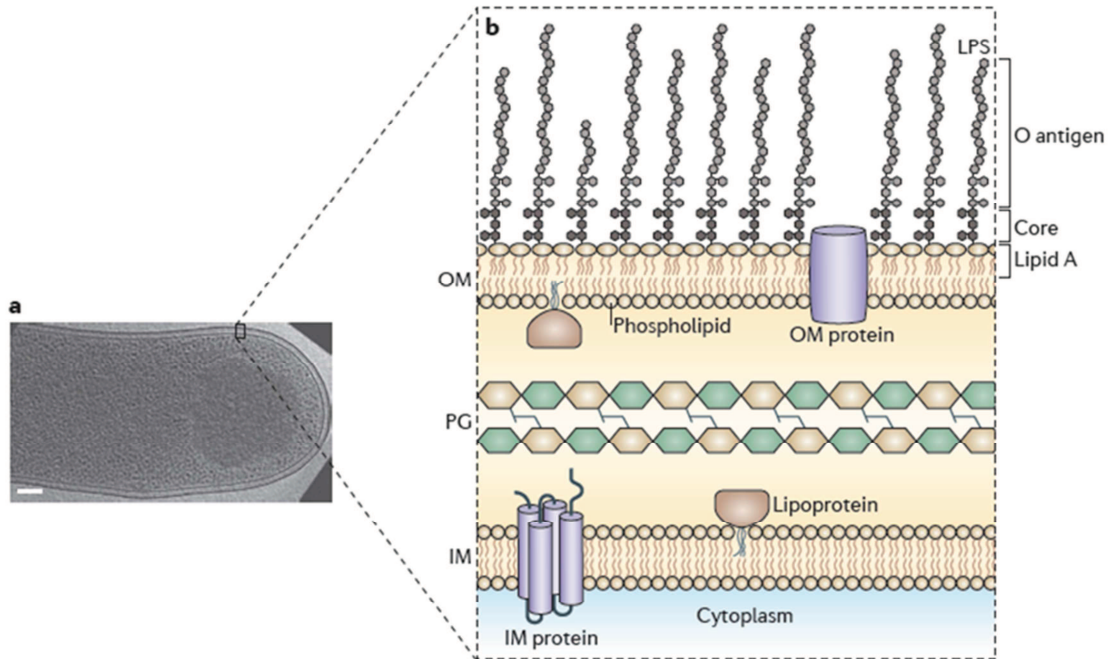
Trois anticorps ont été produits afin de les coupler aux billes magnétiques : FTC129Z, un anticorps IgG2B de souris, FT100R et FT101R des anticorps IgG de rat.

### 3.2 Cible des anticorps

La cible des anticorps est le lipopolysaccharide (LPS) des *Francisella tularensis* subsp. *holarctica* LVS.

#### 3.2.1 Données structurales sur le LPS

Le lipopolysaccharide (LPS) est le composant non protéique de la membrane externe des bactéries à Gram négatif. Comme nous pouvons le voir sur la **Figure 22**, il est constitué de trois parties distinctes assemblées entre elles. Le lipide A est hydrophobe et ancré dans la bicouche lipidique et il est responsable de l'activité endotoxinique de la bactérie. En effet, lors de la lyse bactérienne il y a une libération importante de LPS. La structure du lipide A est fortement conservée parmi l'ensemble des bactéries à Gram négatif. L'antigène O ou antigène somatique (antigène du corps de la bactérie), tourné vers l'extérieur, est l'interface entre la bactérie et le milieu extérieur. C'est un polysaccharide à caractère hydrophile qui est constitué de blocs d'hydrates de carbone, ou encore « unités O », assemblés entre eux par polymérisation. Les unités O sont composées de un à huit résidus et se répètent jusqu'à 50 fois pour former l'antigène O. Les différentes intensités de polymérisation des unités O entraînent la présence, en surface de la bactérie, d'une grande variété de longueurs d'antigènes O. Les unités O peuvent être composées d'un seul et même type d'hydrate de carbone (homopolymère) ou bien de plusieurs types différents (hétéropolymère). Ils sont assemblés de manière linéaire ou ramifiée. Enfin le noyau, ou « core », relie les deux parties précédentes. Il est composé d'hydrates de carbone et de dérivés d'hydrates de carbone et est divisé en deux parties : le noyau interne plutôt hydrophobe et le noyau externe plutôt hydrophile (Szalo et al., 2006). La structure générale du noyau interne est fortement conservée chez les bactéries à Gram négatif alors que le noyau externe est plus variable.



**Figure 22** : Schéma du LPS (Needham and Trent, 2013). **a)** Image d'une cellule d'*Escherichia coli* obtenue par tomographie par cryomicroscopie. Barre d'échelle : 200 nm. **b)** Schéma de l'enveloppe d'une cellule gram-négative montrant les bicouches de la membrane interne (IM) et de la membrane externe (OM) séparées par le périplasma contenant le peptidoglycane (PG). C'est dans la bicouche lipidique de la membrane externe que sont fixés les lipides A du LPS.

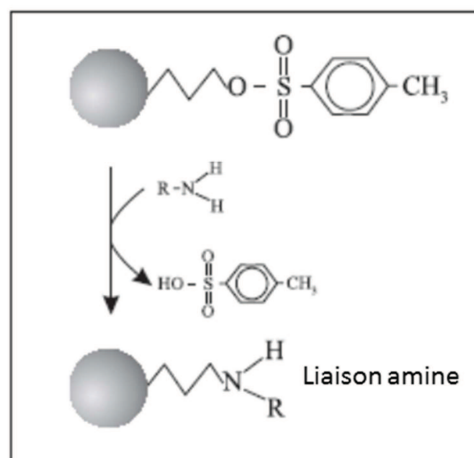
### 3.2.2 Données fonctionnelles sur le LPS

Le LPS joue un rôle très important dans la relation hôte-bactérie avec des implications éventuelles dans la pathogenèse des infections bactériennes. Le rôle du LPS dans la pathogénie est dû essentiellement au lipide A et à l'antigène O. En effet, l'antigène O, présenté vers l'extérieur de la bactérie et représentant la partie immunogénique spécifique du LPS, influence la relation hôte-bactérie à différents niveaux en assurant plusieurs fonctions comme la résistance à l'action du complément, la résistance à la phagocytose, le mimétisme moléculaire et l'adhérence aux tissus de l'hôte (Li et al., 2012). C'est l'antigène O qui présente le plus de variations entre les différentes espèces bactériennes et qui est donc décisif pour la spécificité antigénique du LPS. Par exemple, chez *E. coli* plus de 170 sérogroupes O ont été décrits (Raetz and Whitfield, 2002). Cependant, il est également possible de trouver des espèces bactériennes très éloignées qui possèdent des structures semblables d'unités O tels que chez *Brucella* sp. sous-type A, de *Yersinia enterocolitica* O9 et de *Vibrio cholerae* type *inaba* (Szalo et al., 2006).

#### 4 Les billes magnétiques

Nous avons testé plusieurs technologies de billes magnétiques de manière à trouver les billes magnétiques pour lesquelles le couplage avec l'anticorps et le résultat de l'immunocapture sont les plus efficaces. Les trois technologies décrites ci-dessous ont des billes de l'ordre du  $\mu\text{m}$  et le couplage billes-anticorps s'effectue de manière covalente.

- Dynabeads MyOne® Tosylactivated™ (Invitrogen). Ce sont des billes super-paramagnétiques en polystyrène recouvertes d'une couche de polyuréthane de 1  $\mu\text{m}$  de diamètre. Les groupements hydroxyles en surface sont activés par du chlorure de tosyloxy (chlorure de 4-toluènesulfonyl). L'ester (tosylate) obtenu peut réagir de manière covalente avec les groupements amines des anticorps comme le montre le schéma réactionnel de la **Figure 23**.



**Figure 23** : Schéma réactionnel entre un sulphonyl ester et le groupement amine d'un anticorps. Dynabeads® MyOne™ Tosylactivated (Invitrogen).

- MasterBeads Carboxylic Acid 0215 (Ademtech). Ce sont des billes super-paramagnétiques de 500 nm de diamètre. Le centre de la bille est magnétique et recouvert d'une enveloppe en polymère hydrophile hautement réticulé. La surface est fonctionnalisée avec des acides carboxyliques. Ceux-ci sont activés avec de l'EDC (1-ethyl-3-(3-diméthylaminopropyl) de manière à former des groupements O-acyluré réactifs avec les amines primaires.
- Activ-Masterbeads 0265 (Ademtech). Ce sont des billes super-paramagnétiques de 510 nm de diamètre. Le centre de la bille est magnétique et recouvert d'une enveloppe en polymère hydrophile hautement réticulé. La fonctionnalisation de la surface permet de fixer les protéines de manière covalente sans activation préalable mais elle n'est pas décrite pour des raisons de secret de fabrication.

## 5 Méthodes analytiques pour l'évaluation des résultats

Pour valider les différentes étapes du protocole d'immunocapture, plusieurs outils ont été utilisés : la mesure d'absorbance, le test ELISA et la spectrométrie de masse MALDI-TOF, l'outil final de détection de cette méthode d'immunocapture.

### 5.1 Mesure d'absorbance

La première étape de ce protocole d'immunocapture est le couplage des anticorps à la surface des billes. Afin d'évaluer la capacité de fixation des billes et le rendement de fixation des billes, nous avons choisi d'utiliser la mesure de densité optique à 280 nm. Les groupements aromatiques des tyrosines et des tryptophanes absorbent à cette longueur d'onde, ce qui permet de mesurer la concentration en protéines d'une solution. La quantité d'anticorps fixés par les différents types de billes a été évaluée par mesure de densité optique à 280 nm (Carry 50, Varian) en mesurant l'absorbance de la solution d'anticorps avant le couplage et l'absorbance du surnageant après le couplage. Le mode de calcul de la concentration est détaillé dans l'Équation 2.

**Équation 2** : Calcul de la concentration en anticorps basé sur la loi de Beer-Lambert.

$$C = \frac{A_{280nm}}{l \times \epsilon_{280nm}}$$

- $C$  : Concentration en g.L<sup>-1</sup>
- $A_{280nm}$  : Absorbance à 280 nm
- $l$  : Longueur de la cuve en cm
- $\epsilon_{280nm}$  : Coefficient d'extinction molaire. Ici  $\epsilon_{280nm}=1,4 \text{ L.g}^{-1}.\text{cm}^{-1}$

### 5.2 Test ELISA

Pour établir la quantité de bactéries présentes dans les solutions à chaque étape du protocole, nous avons réalisé des tests ELISA avec des courbes étalons. Cette méthode est choisie pour sa facilité de mise en œuvre et pour sa sensibilité. Nous pouvons ainsi déterminer des concentrations jusqu'à 10<sup>4</sup> CFU/ml. Les anticorps mis en jeu dans ce test ELISA sont FT100R et FT101R. Tout d'abord, les microplaques de 96 puits sont saturées par l'anticorps FT101R. L'anticorps biotinylé (FT100R) est disposé dans tous les puits avant d'ajouter les échantillons. Une gamme étalon est préparée à partir d'une solution de *Francisella tularensis* subsp. *holarctica* LVS. Cette gamme est déposée en duplicata dans les puits de la microplaque pour une incubation d'une nuit à 4°C. En parallèle, les solutions à doser sont diluées au 50<sup>ème</sup>, 500<sup>ème</sup> et 5 000<sup>ème</sup> avant d'être déposées en triplicata sur la microplaque pour l'incubation. Ces dilutions sont données à titre indicatif, selon la concentration de l'échantillon elles peuvent être modifiées de manière à ce que l'échantillon dilué soit dans la gamme de détection du test ELISA. Au terme de l'incubation, les bactéries présentes dans l'échantillon sont fixées à la fois à l'anticorps FT101R dont la plaque est saturée et à l'anticorps FT100R couplé à une biotine. Après l'incubation, la microplaque est lavée et une solution de streptavidine-peroxydase est distribuée dans les puits avant 30 minutes d'incubation à température ambiante. Comme la streptavidine possède une forte

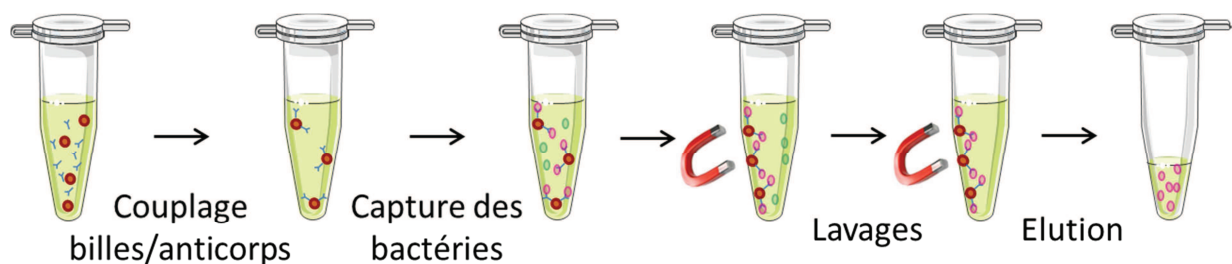
affinité avec la biotine, les couples streptavidine-peroxydase vont se fixer sur les biotines. La plaque est de nouveau lavée et une solution de TMB (3,3',5,5'-tétraméthylbenzidine) est distribuée dans les puits. Le TMB va révéler la peroxydase après une incubation à l'abri de la lumière en devenant bleu. La dernière étape consiste à arrêter la réaction de la peroxydase en ajoutant de l'acide sulfurique, le produit formé par le TMB sous l'action de la peroxydase va alors devenir jaune. La lecture de la densité optique (DO) se fait à 450 nm avec un spectrophotomètre (Multiskan® Spectrum, Thermo Scientific). A partir des résultats obtenus pour la gamme de concentration il est possible de déterminer la concentration des échantillons.

### 5.3 Spectrométrie de masse MALDI-TOF

L'échantillon après immunocapture a pour objectif d'être analysé par spectrométrie de masse MALDI-TOF (Biflex IV, Bruker). Il est donc nécessaire que la qualité des spectres soit suffisante pour faire de la détection. L'observation des spectres (bruit de fond, résolution et intensité des pics, reproductibilité) nous a parfois permis de choisir entre plusieurs protocoles possibles.

## 6 Optimisation du protocole d'immunocapture

La **Figure 24** représente les différentes étapes ayant lieu lors du protocole d'immunocapture et sur lesquelles nous avons réalisé des optimisations. La première étape consiste à coupler les anticorps aux billes magnétiques. Ensuite, les billes magnétiques couplées vont permettre de réaliser l'immunocapture à proprement parler : c'est-à-dire la capture des bactéries durant un temps d'incubation, les différents lavages pour purifier les bactéries et enfin l'élution pour libérer les bactéries des billes. Les optimisations sont détaillées dans les prochains paragraphes.



**Figure 24** : Schéma général du protocole d'immunocapture.

### 6.1 Choix de l'anticorps

Les rendements de couplage ont été mesurés pour les trois anticorps FTC129Z, FT100R et FT101R sur les Dynabeads MyOne® Tosylactivated™ pour un couplage de 40 µg d'anticorps par mg de billes. Pour FTC129Z, le rendement de couplage moyen est de  $72 \pm 7$  %. Pour FT100R et FT101R, le rendement moyen est de  $95 \pm 1$  %. Avec un meilleur rendement de



couplage et pour des raisons de disponibilité des anticorps, c'est l'anticorps FT100R qui a été choisi pour l'ensemble des expériences du banc d'essai.

## 6.2 Couplage des anticorps

Pour les billes Dynabeads MyOne® Tosylactivated™, MasterBeads Carboxylic Acid 0215 et Activ-Masterbeads 0265, nous avons fixé des quantités variables d'anticorps anti-bactéries sur une quantité fixe de billes (4 mg): 5, 15, 25, 50 µg d'Ac /mg de billes. La quantité d'anticorps fixés par les différents types de billes a été évaluée par mesure de densité optique (DO) du surnageant de couplage, permettant ainsi de calculer le rendement de couplage.

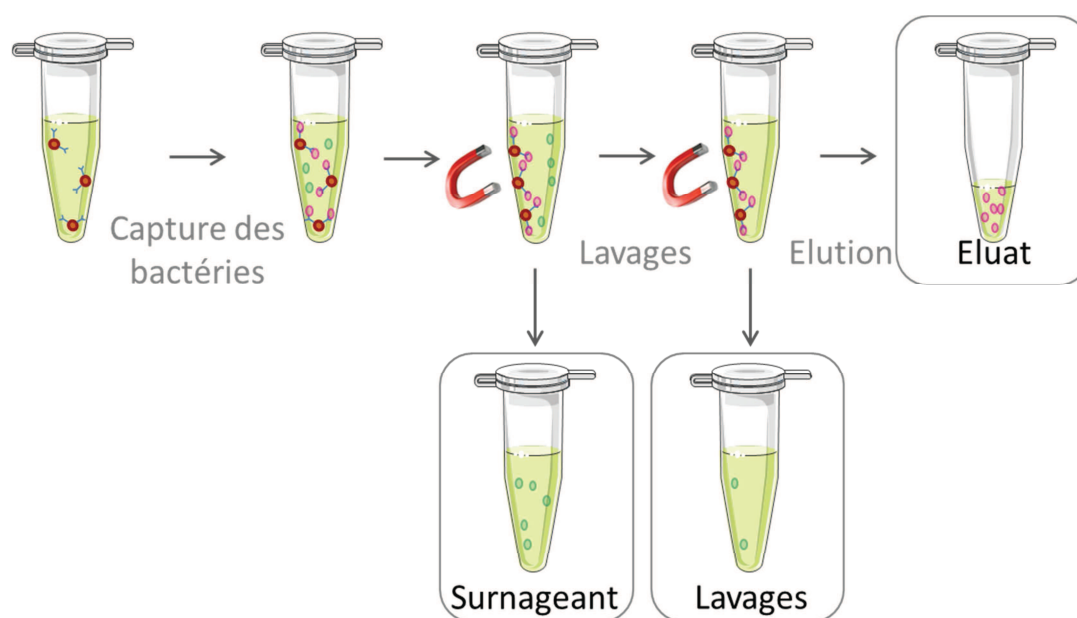
Pour cela, la mesure de DO à 280 nm a été effectuée sur les surnageants de couplage, ainsi que sur un échantillon d'anticorps avant mise en contact des billes, comme référence. Tel qu'indiqué dans le **Tableau 8**, les deux types de billes Ademtech montrent un très faible rendement de couplage d'anticorps. D'une part, les billes MasterBeads Carboxylic Acid 0215 ne couplent aucun anticorps. D'autre part, pour les billes Activ-MasterBeads 0265, le taux de rendement est de 100, 32, 26 et 24% pour 5, 15, 25 et 50 µg d'Ac/mg de billes respectivement. Les billes Dynabeads MyOne® Tosylactivated™ ont un rendement de fixation beaucoup plus intéressant, comme le montre le **Tableau 8**, avec environ 85 % de rendement pour les 3 plus faibles concentrations d'anticorps, et 60 % pour 50 µg d'Ac/mg de billes, ce qui pourrait être expliqué par une saturation des billes par les anticorps. Au vu de ces résultats, nous avons choisi de tester 6 tubes de billes (Activ-MasterBeads 0265 à 25 et 50 µg d'Ac/mg de billes et Dynabeads MyOne® Tosylactivated™ à 5, 15, 25 et 50 µg d'Ac/mg de billes) pour l'immunocapture de bactéries actives et de faire des tests ELISA de chaque solution (surnageant, lavage, éluat) afin de connaître leur concentration exacte en bactéries, puis les spectres MALDI-TOF des éluats afin d'en évaluer la qualité.

**Tableau 8** : Rendement de couplage des différents types de billes. Les concentrations en anticorps des surnageants de couplage sont calculées à partir des mesures de densité optique à 280 nm. Les lignes en gras représentent les billes conservées pour l'immunocapture de bactéries.

Billes	Quantité d'Ac (µg) mis en présence par mg de billes	Couplage déduit de la concentration des surnageants (µg d'Ac/mg de billes)	Rendement de couplage
Dynabeads MyOne	<b>5 µg/mg</b>	<b>4 µg/mg</b>	<b>80 %</b>
	<b>15 µg/mg</b>	<b>13 µg/mg</b>	<b>88 %</b>
	<b>25 µg/mg</b>	<b>22 µg/mg</b>	<b>90 %</b>
	<b>50 µg/mg</b>	<b>29 µg/mg</b>	<b>58 %</b>
Activ-MasterBeads	5 µg/mg	5 µg/mg	100 %
	15 µg/mg	5 µg/mg	32 %
	<b>25 µg/mg</b>	<b>7 µg/mg</b>	<b>26 %</b>
	<b>50 µg/mg</b>	<b>12 µg/mg</b>	<b>24 %</b>
MasterBeads Carboxylic Acid 0215	5 µg/mg	0 µg/mg	0 %
	15 µg/mg	0 µg/mg	0 %
	25 µg/mg	0 µg/mg	0 %
	50 µg/mg	0 µg/mg	0 %

### 6.2.1 Test ELISA

Tout d'abord une gamme de dilution de *Francisella tularensis* subsp. *holarctica* LVS a été préparée pour le test ELISA. Pour cela nous avons fait une dilution en cascade des bactéries de  $10^9$  à  $10^3$  CFU/ml suivie d'une inactivation à la chaleur. Les échantillons sont testés dans un premier temps avec des dilutions au 50<sup>ème</sup>, 500<sup>ème</sup> et 5 000<sup>ème</sup>. L'analyse de la gamme montre qu'il y a saturation pour les concentrations les plus importantes (de  $10^9$  à  $10^7$  CFU/ml). Nous utilisons donc uniquement les concentrations de  $10^6$  à  $10^4$  CFU/ml pour calculer les concentrations de bactéries présentes dans chaque échantillon. Pour les échantillons trop concentrés, nous les avons testés lors d'une seconde analyse dilués au 50 000<sup>ème</sup>, 500 000<sup>ème</sup> et 5 000 000<sup>ème</sup>. Nous avons également déterminé la concentration exacte de notre échantillon initial grâce à la gamme. La valeur calculée est de  $5.10^8$  CFU/ml au lieu des  $10^9$  CFU/ml attendus. Les différentes étapes du protocole et les échantillons testés sont présentés dans la **Figure 25**.



**Figure 25** : Récapitulatif des différentes étapes de l'immunocapture. Le surnageant est la solution qui est éliminée après l'incubation des billes avec l'échantillon à enrichir. Les lavages sont les trois solutions qui sont éliminées après les étapes de lavages. L'éluat est le résultat de l'immunocapture.

Le **Tableau 9** recense les résultats des tests ELISA qui sont détaillés ci-dessous. Dans les surnageants (volume : 1 ml) des billes Dynabeads MyOne® Tosylactivated™ couplées à 15, 25 et 50 µg d'Ac/mg de billes, la concentration est comprise entre  $3.10^7$  et  $6.10^7$  CFU/ml, soit environ 10 % de la quantité de bactéries mise en jeu lors de l'immunocapture. A 5 µg d'Ac/mg de billes, la concentration est de  $(4,3 \pm 0,2).10^8$  CFU/ml. Pour les billes Ademtech, les concentrations peuvent être estimées inférieures à  $10^6$  CFU/ml car la valeur du signal est indiscernable du bruit de fond. Dans les lavages (volume : 1,2 ml), les concentrations de bactéries présentes pour les billes Dynabeads MyOne® Tosylactivated™ couplées à 15, 25 et 50 µg d'Ac/mg de billes sont de l'ordre de  $3.10^7$  CFU/ml, soit moins de 10 % des bactéries introduites au départ. Pour les billes Ademtech, la concentration peut être estimée

inférieure à  $10^6$  CFU/ml. Les éluats présentent des concentrations beaucoup plus intéressantes. En effet, pour les billes Dynabeads MyOne® Tosylactivated™ couplées à 15, 25 et 50  $\mu\text{g}$  d'Ac/mg de billes, les quantités de bactéries éluées sont proches de la quantité introduite (6,4 et  $4 \cdot 10^8$  bactéries). A 5  $\mu\text{g}$  d'Ac/mg de billes Dynabeads MyOne® Tosylactivated™ et pour les billes Ademtech, la quantité observée varie entre  $2 \cdot 10^8$  et  $3 \cdot 10^8$  CFU ce qui laisse penser qu'une partie des bactéries n'est pas fixée (Dynabeads MyOne® Tosylactivated™ couplées à 5  $\mu\text{g}$  d'Ac/mg de billes) ou qu'une partie des bactéries n'est pas éluée (billes Ademtech).

**Tableau 9** : Résultats des tests ELISA pour déterminer la concentration de *Francisella tularensis* subsp. *holarctica* LVS dans les solutions aux différentes étapes de l'immunocapture.

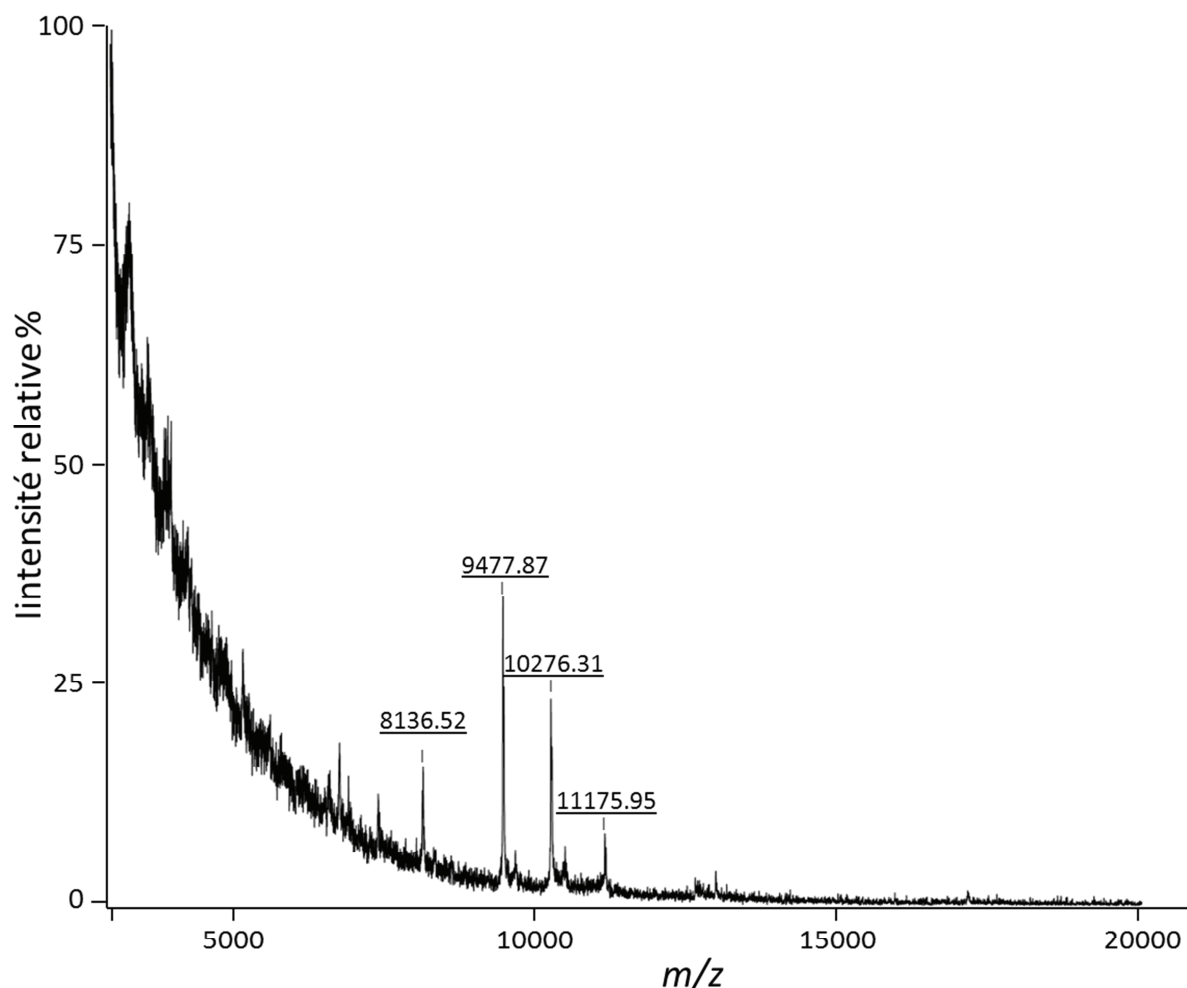
Billes	Quantité d'Ac ( $\mu\text{g}$ ) mis en présence par mg de billes	Concentration en CFU/ml		
		Surnageant	Lavages	Eluat
Dynabeads MyOne	5	$(4,3 \pm 0,2) \cdot 10^8$	$(1,3 \pm 0,4) \cdot 10^8$	$(1,8 \pm 0,7) \cdot 10^9$
	15	$(3,0 \pm 0,8) \cdot 10^7$	$(3,8 \pm 0,3) \cdot 10^7$	$(6,2 \pm 0,2) \cdot 10^9$
	25	$(5,6 \pm 1,5) \cdot 10^7$	$(2,7 \pm 0,8) \cdot 10^7$	$(4,2 \pm 0,0) \cdot 10^9$
	50	$(6,1 \pm 1,5) \cdot 10^7$	$(2,4 \pm 0,9) \cdot 10^7$	$(3,6 \pm 0,4) \cdot 10^9$
Activ- MasterBeads	25	$< 10^6$	$< 10^6$	$(3,2 \pm 0,1) \cdot 10^9$
	50	$< 10^6$	$< 10^6$	$(2,1 \pm 0,5) \cdot 10^9$

Le test ELISA nous permet de montrer qu'il y a bien eu capture des bactéries par les billes et que, selon les billes, nous pouvons capturer et éluer presque la totalité des bactéries introduites à l'état initial. Les billes donnant les meilleurs résultats sont les Dynabeads MyOne® Tosylactivated™ couplées à 15, 25 et 50  $\mu\text{g}$  d'Ac/mg de billes. Nous constatons que la corrélation entre quantité d'anticorps couplés et concentration de bactéries dans l'éluat n'est pas directe. En effet, avec 50  $\mu\text{g}$  d'Ac/mg de billes il n'y a pas plus de bactéries éluées que pour les billes à 15 ou 25  $\mu\text{g}$  d'Ac/mg de billes. Il faut toutefois garder à l'esprit que ces concentrations ne sont que des approximations comme le montre les écarts-types parfois importants dans le **Tableau 9**. En effet, la saturation aux concentrations importantes ne permet pas d'avoir beaucoup de points utilisables pour établir la gamme, ce qui limite la plage d'utilisation de celle-ci.

### 6.2.2 Analyse par spectrométrie de masse MALDI-TOF

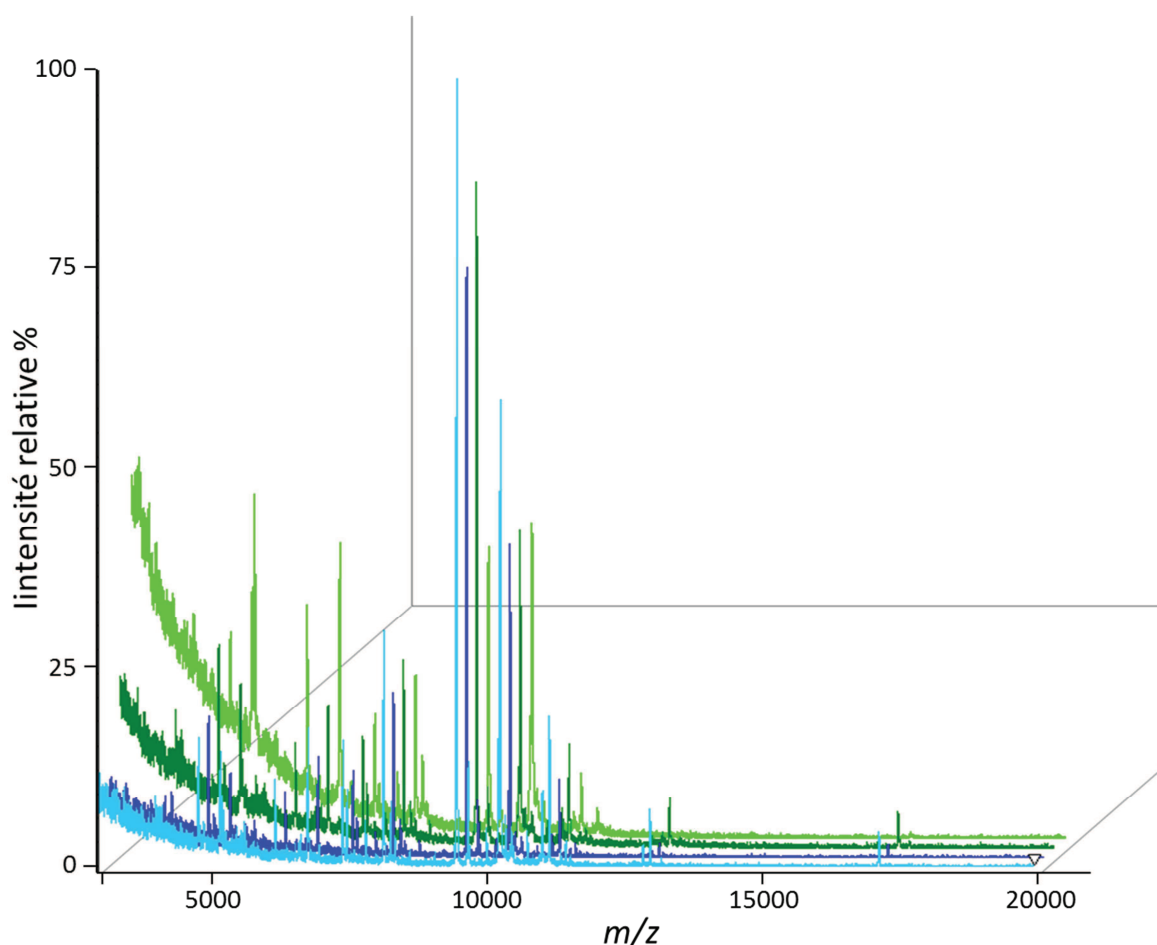
Les échantillons correspondants aux éluats sont déposés sur cible MALDI dans les conditions suivantes : un volume de 1  $\mu\text{l}$  de l'échantillon auquel est ajouté un volume de 1  $\mu\text{l}$  d'acide sinapinique (matrice SA). Pour les spectres bien résolus, une nouvelle analyse est menée en ajoutant un calibrant interne (Protein Calibration Standard I, Bruker). Les échantillons sont analysés sur un MALDI-TOF Biflex IV (Bruker) dans la gamme de  $m/z$  de 3 000 à 20 000.

Les spectres obtenus avec les bactéries isolées à partir des billes *Activ-MasterBeads* ne sont pas de bonne qualité comme le montre la **Figure 26**. Le spectre comporte peu de pics, de faible intensité, et un bruit de fond important. Seuls les éluats des billes *Dynabeads MyOne® Tosylactivated™* ont donné des spectres permettant de réaliser l'identification comme le montre la **Figure 27**.



**Figure 26** : Spectre MALDI-TOF de l'éluat des billes *Activ-MasterBeads* couplées à 50 µg d'Ac/mg de billes acquis sur la gamme de  $m/z$  de 3 000 à 20 000.

L'intensité des signaux  $m/z$ , la résolution et le bruit de fond sur les spectres des éluats des billes *Dynabeads MyOne® Tosylactivated™* varie selon la concentration initiale d'anticorps couplés aux billes. Nous pouvons constater sur la **Figure 27** une nette diminution du bruit de fond quand les quantités d'anticorps couplés augmentent, avec les meilleurs spectres à 25 et 50 µg d'Ac/mg de billes.



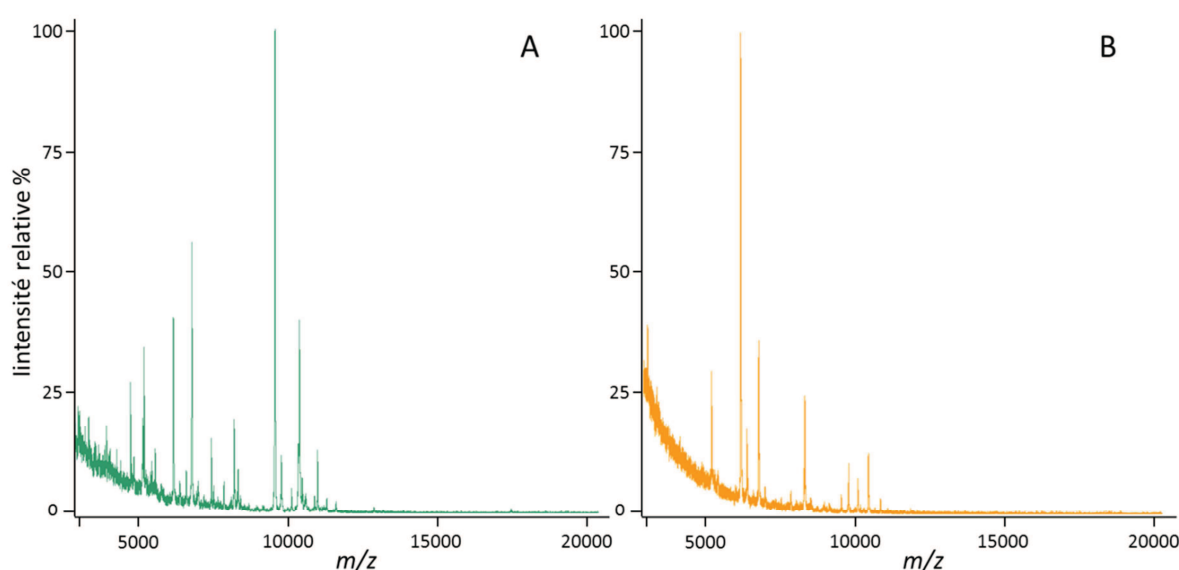
**Figure 27** : Spectres MALDI-TOF des éluats des billes Dynabeads MyOne® Tosylactivated™ couplées à 5, 15, 25 et 50 µg d'Ac/mg de billes acquis sur la gamme de  $m/z$  de 3 000 à 20 000. Vert clair : 5 µg d'Ac/mg de billes ; vert foncé : 15 µg d'Ac/mg de billes ; bleu foncé : 25 µg d'Ac/mg de billes ; bleu clair : 50 µg d'Ac/mg de billes.

D'après les résultats obtenus avec les tests ELISA et les spectres MALDI-TOF, les billes Dynabeads MyOne® Tosylactivated™ couplées à 25 et 50 µg d'Ac/mg de billes permettent d'enrichir l'échantillon en bactéries et la qualité des spectres de leur éluat (nombre de pics, résolution et intensité des pics) permet de faire une identification. Les spectres de ces deux concentrations de couplages sont équivalents. Pour cette raison et par souci d'économie, nous avons choisi les billes Dynabeads MyOne® Tosylactivated™ avec un couplage à 25 µg d'Ac/mg de billes.

### 6.3 Cible de l'immunocapture

Afin d'éviter de travailler dans les conditions d'un laboratoire de sécurité biologique de niveau 2 pour la mise en œuvre de l'enrichissement des bactéries, ces expériences ont été réalisées à la fois avec des bactéries vivantes, issues de cultures fraîchement préparée et des bactéries inactivées à la chaleur et stockées au froid en tampon salin, 65 % glycérol (GSB). Les résultats d'enrichissement des bactéries vivantes et des bactéries inactivées à la chaleur ont ensuite été comparés. Le GSB est utilisé pour assurer l'intégrité des bactéries lors de la congélation et de la décongélation. Seulement malgré les lavages lors de l'immunocapture,

la totalité du glycérol ne parvient pas à être éliminé. Le glycérol résiduel pose un problème au moment des dépôts sur la cible MALDI. Il ne s'évapore pas, et de ce fait empêche la cristallisation des dépôts rendant l'analyse MALDI-TOF impossible. Dans le cas des bactéries inactivées à la chaleur et stockées en PBS (tampon phosphate salin), les spectres MALDI-TOF des éluats sont très différents en termes de profils protéiques par rapports aux spectres d'éluats d'immunocaptures sur des bactéries vivantes. En effet, les spectres des éluats issus d'immunocaptures sur des bactéries inactivées présentent peu de pics comme nous pouvons l'observer sur la **Figure 28**. Cela peut s'expliquer par le fait que les bactéries, endommagées par l'inactivation à la chaleur, ont perdu leur contenu protéique au moment des lavages réalisés pendant l'immunocapture ciblée sur le LPS. Les spectres montrent donc moins d'informations et le profil protéique est différent. Dans ce cas, les protéines qui apparaissent sur les spectres sont certainement des protéines liées au LPS.



**Figure 28** : Spectres MALDI-TOF des éluats d'immunocapture, acquis dans la gamme de  $m/z$  de 3 000 à 20 000. **A** : éluat d'immunocapture de bactéries vivantes. **B** : éluat d'immunocapture de bactéries inactives.

En conclusion, comme nous souhaitons avoir le maximum d'informations possible sur les spectres pour identifier les bactéries sans ambiguïtés, nous avons choisi de travailler sur les bactéries actives. Cela entraîne donc des contraintes supplémentaires : travail en L2, travail sous PSM, inactivation et validation de l'inactivation avant de pouvoir analyser le résultat de l'immunocapture.

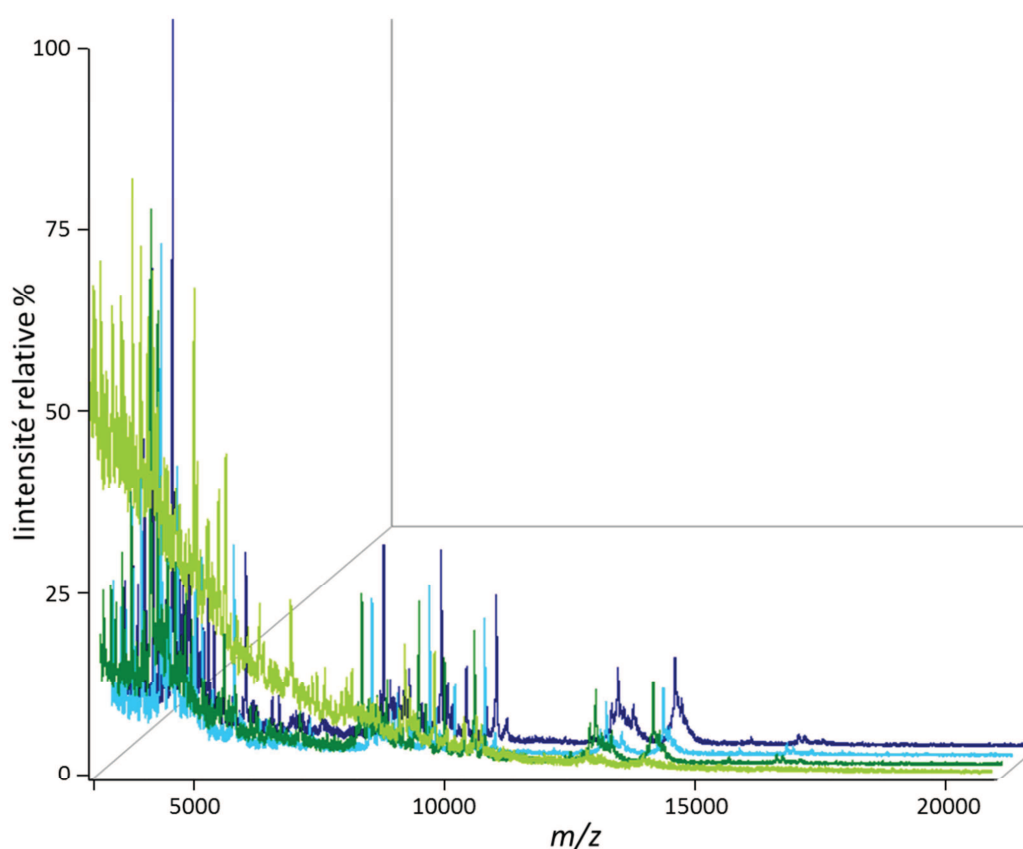
#### 6.4 Inactivation

Comme nous avons choisi de faire l'immunocapture sur les bactéries actives, il est nécessaire d'inactiver les échantillons avant l'analyse par spectrométrie de masse. Lors des premières expérimentations sur les bactéries actives, nous nous sommes aperçus que les spectres de tous les éluats, même ceux des témoins négatifs, étaient semblables et cela quelle que soit la concentration initiale, comme nous pouvons le voir sur la **Figure 29**, la

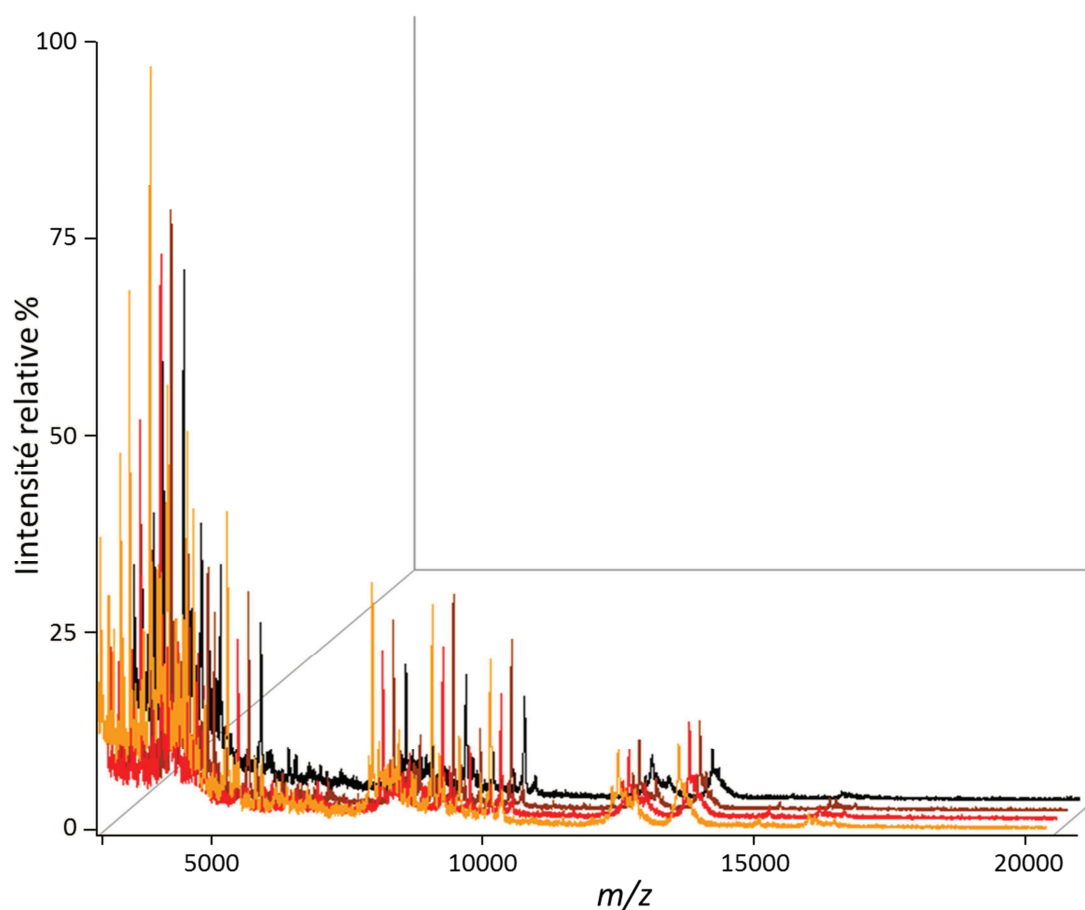
**Figure 30** et la **Figure 31**. Pour constituer les témoins négatifs, deux types d'échantillons ont été mis en œuvre :

- Les premiers ont été préparés à partir de solutions de PBS stérile ayant suivi toutes les étapes de l'immunocapture.
- Les seconds ont été préparés à partir de solutions de *Francisella tularensis* subsp. *holarctica* LVS à différents niveaux de concentrations dans du PBS, avec des billes couplées à des anticorps anti-TMPP, aspécifiques de la bactérie cible.

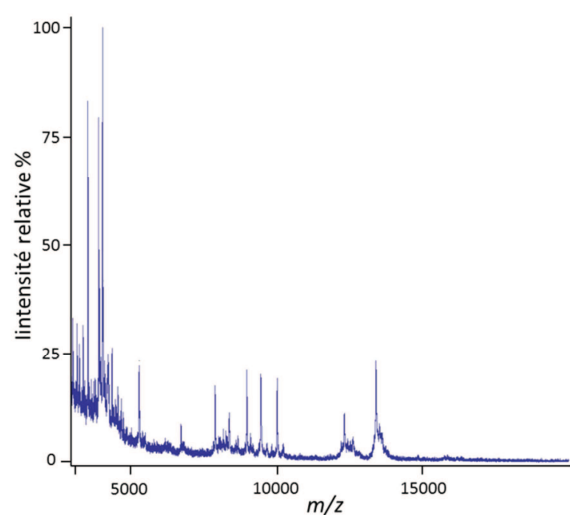
Sur la **Figure 32**, qui représente l'analyse de l'éluat d'une immunocapture sur du PBS stérile et sans chauffage, nous n'observons, cette fois-ci, aucun pic. En effet, le protocole initial du constructeur proposait l'utilisation de PBS BSA 0,1 % pour faire le couplage des billes et pour les lavages lors de l'immunocapture. Nous avons testé l'immunocapture sans mettre de BSA dans le PBS et cette fois-ci les nombreux pics observés sur le spectre en **Figure 31** n'apparaissaient plus. Ces pics étaient le résultat de l'action du chauffage sur le milieu acide, ce qui provoquait une hydrolyse acide de la BSA présente en quantité importante dans le milieu. De ce fait, pour éviter l'apparition du signal dû à l'hydrolyse de la BSA qui masquait complètement les signaux spécifiques des échantillons, nous avons décidé de supprimer la BSA de tous les tampons intervenant dans l'expérience et de neutraliser le pH de l'éluat avant l'inactivation avec de l'hydroxyde d'ammonium de manière à éviter l'hydrolyse acide.



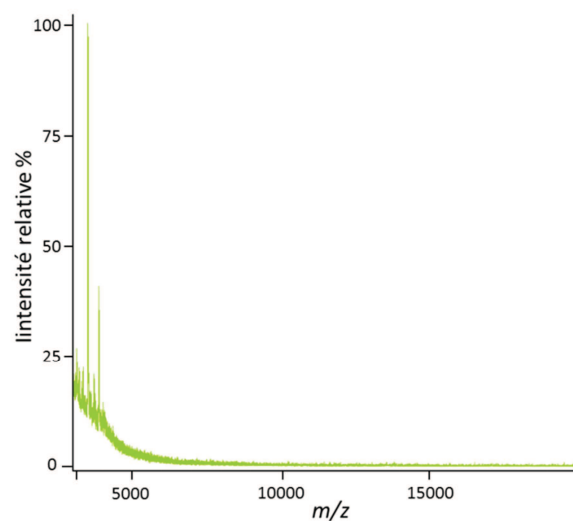
**Figure 29** : Spectres MALDI-TOF, éluats des billes couplées aux anticorps FT100R de  $10^9$  CFU/ml (vert clair) à  $10^6$  CFU/ml (bleu foncé).



**Figure 30** : Spectres MALDI-TOF, éluats des billes couplées aux anticorps TMPP de  $10^9$  CFU/ml (orange) à  $10^6$  CFU/ml (noir).



**Figure 31** : Spectre MALDI-TOF, éluat de l'échantillon sans bactéries, chauffé 1h à 90°C.



**Figure 32** : Spectre MALDI-TOF, éluat de l'échantillon sans bactéries et sans chauffage.



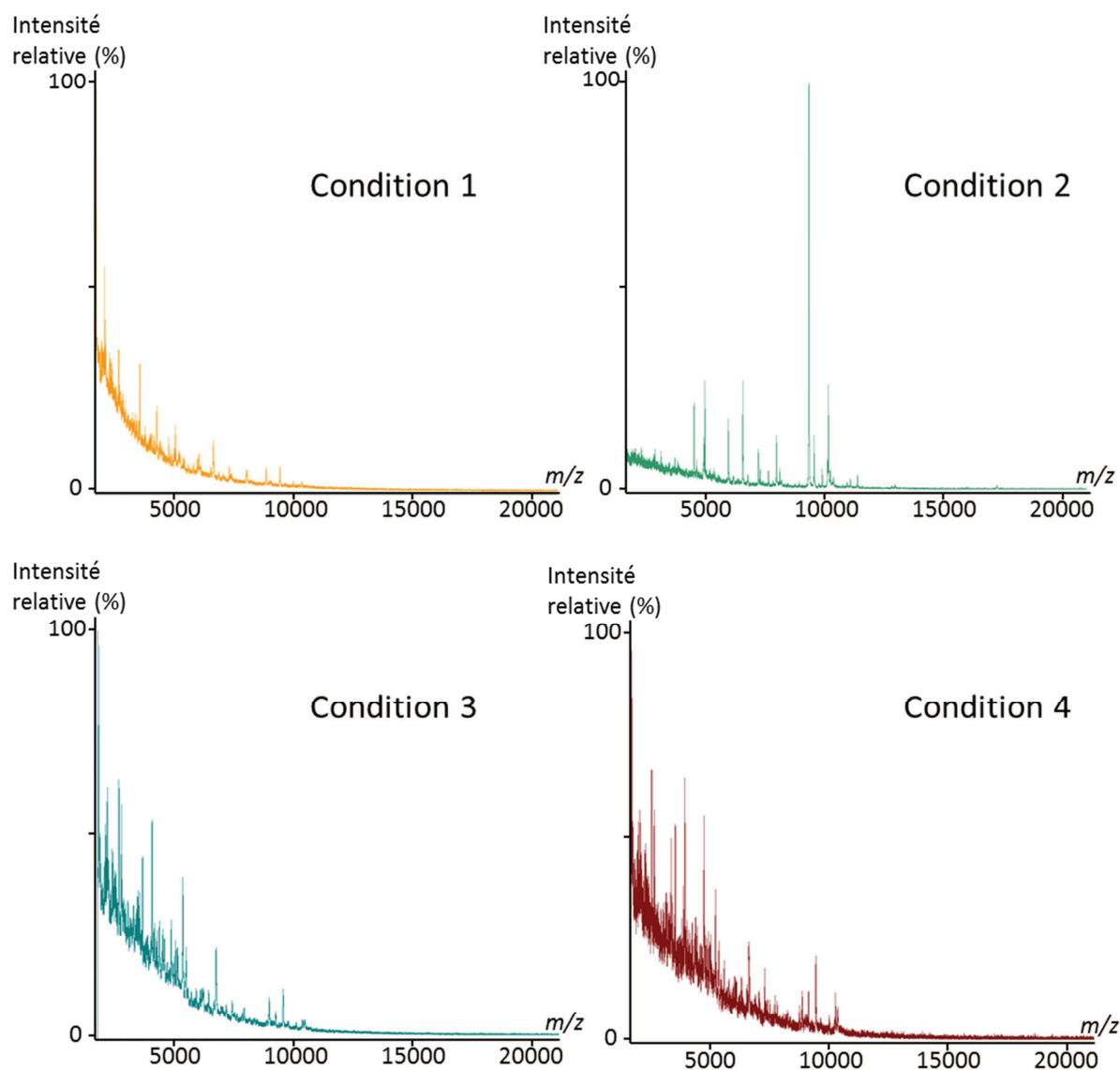
## 6.5 Lavages

Les étapes de lavage ont aussi été optimisées de manière à avoir les spectres de meilleure qualité possible. En effet, même si le MALDI-TOF est tolérant envers les sels, le PBS dans lequel se trouve les bactéries est une solution concentrée en sels ce qui peut entraîner une altération des spectres. C'est pourquoi nous avons expérimenté la possibilité de modifier la série de lavages en PBS par des lavages en eau. Les 4 scénarii de lavages possibles sont détaillés dans le **Tableau 10**.

**Tableau 10** : Répartition des lavages en H<sub>2</sub>O et en PBS selon les tubes.

Condition	Lavage n°1	Lavage n°2	Lavage n°3
<b>1</b>	PBS	PBS	PBS
<b>2</b>	PBS	PBS	H <sub>2</sub> O
<b>3</b>	PBS	H <sub>2</sub> O	H <sub>2</sub> O
<b>4</b>	H <sub>2</sub> O	H <sub>2</sub> O	H <sub>2</sub> O

Pendant l'expérience nous constatons que les billes ayant subi un premier lavage en eau s'agrégent fortement. De plus, plus l'échantillon a subi de lavages en eau, moins l'éluat est trouble, c'est-à-dire qu'il semble y avoir moins de bactéries dans l'éluat. Les spectres MALDI-TOF présenté dans la **Figure 33** sont obtenus en déposant sur une cible 1 µl d'échantillon et 1 µl de matrice SA.



**Figure 33** : Spectres MALDI-TOF des immunocaptures selon les différentes conditions de lavages. En jaune : condition 1, PBS/PBS/PBS. En vert : condition 2, PBS/PBS/H<sub>2</sub>O. En bleu : condition 3, PBS/H<sub>2</sub>O/H<sub>2</sub>O. En rouge : condition 4, H<sub>2</sub>O /H<sub>2</sub>O/H<sub>2</sub>O.

Les quatre conditions montrent d'importantes différences. Faire le dernier lavage en eau (condition 2) augmente le rapport signal sur bruit. En revanche, lorsque les deux derniers lavages sont avec de l'eau (condition 3) ou tous les lavages sont en eau (condition 4), la perte de signal est très importante et le bruit de fond augmente. Ceci est probablement dû à la lyse des cellules en conditions hypotoniques.

Nous avons donc choisi de faire la série de lavages suivante : 1<sup>er</sup> lavage en PBS, 2<sup>ème</sup> lavage en PBS et 3<sup>ème</sup> lavage en eau.

## 6.6 Elution

### 6.6.1 Tampon

L'élution des bactéries se produit lorsque le pH du milieu passe en condition acide. Le protocole du fournisseur des billes proposait une élution par de l'acide chlorhydrique (HCl) à 50 mM. Nous avons également choisi d'optimiser cette étape du protocole en choisissant des tampons permettant de mettre les billes en condition acide et compatible avec la spectrométrie de masse. Nous avons choisi une élution avec du TFA à 0,2%.

### 6.6.2 Volume

Nous avons aussi essayé d'optimiser le volume d'élution et le volume de bille de manière à avoir la solution de bactéries la plus concentrée possible au moment de l'élution. Nous avons testé les paramètres suivants sur un volume de 1 ml de bactéries à une concentration initiale de  $10^9$  CFU/ml : des volumes de billes de 8  $\mu$ l et 80  $\mu$ l et des volumes d'élution de 10  $\mu$ l, 20  $\mu$ l et 100  $\mu$ l, comme détaillé dans le **Tableau 11**.

**Tableau 11** : Optimisation de la quantité de billes et du volume d'élution. Les résultats du test ELISA donnent les concentrations en bactéries dans le surnageant et dans l'éluat pour chaque essai.

Anticorps	Volume de billes	Volume d'élution	Enrichissement théorique possible	Concentration finale attendue (CFU/ml)	Résultats du test ELISA	
					Concentration surnageant (CFU/ml)	Concentration éluat (CFU/ml)
Anti-His	80 $\mu$ l	100 $\mu$ l	0	0	$(1,0 \pm 0,4) \cdot 10^9$	$< 10^6$
	8 $\mu$ l	10 $\mu$ l	X 100	$10^{11}$	$(1,2 \pm 0,1) \cdot 10^9$	$(6,3 \pm 0,6) \cdot 10^8$
FT100R	8 $\mu$ l	20 $\mu$ l	X 50	$5 \cdot 10^{10}$	$(1,6 \pm 0,6) \cdot 10^9$	$(1,7 \pm 0,3) \cdot 10^9$
	8 $\mu$ l	100 $\mu$ l	X 10	$10^{10}$	$(1,7 \pm 0,6) \cdot 10^9$	$(3,1 \pm 0,8) \cdot 10^9$
	80 $\mu$ l	20 $\mu$ l	X 50	$5 \cdot 10^{10}$	$(2,2 \pm 1,3) \cdot 10^8$	$(6,9 \pm 1,4) \cdot 10^8$
	80 $\mu$ l	100 $\mu$ l	X 10	$10^{10}$	$(2,2 \pm 1,1) \cdot 10^8$	$(4,6 \pm 1,6) \cdot 10^9$

Nous avons réalisé des tests ELISA sur les surnageants d'incubation et sur les éluats de manière à évaluer les quantités de bactéries fixées par les billes et les concentrations des éluats. L'anticorps Anti-His est un anticorps qui cible les étiquettes 6-histidines. Ici il nous a permis de faire un témoin négatif avec des anticorps non-spécifiques à *Francisella tularensis* subsp. *holarctica* LVS.

Le test ELISA confirme bien qu'aucune fixation de *Francisella tularensis* subsp. *holarctica* LVS aspécifique n'a lieu sur les anticorps du témoin négatif. Dans ce cas, la concentration du surnageant d'incubation avec les billes est équivalente à celle de l'échantillon avant l'immunocapture ( $\approx 10^9$  CFU/ml). Dans les tubes contenant 8  $\mu$ l de billes couplées à l'anticorps spécifique à *Francisella tularensis* subsp. *holarctica* LVS, la concentration du surnageant est aussi de l'ordre de  $10^9$  CFU/ml comme le montre les résultats du test ELISA dans le **Tableau 11**. En revanche, dans les tubes contenant 80  $\mu$ l de billes couplées à l'anticorps spécifique, la concentration du surnageant est de l'ordre de  $10^8$  CFU/ml. L'élution des 80  $\mu$ l de billes dans 20  $\mu$ l de tampon montre une concentration en bactéries moins

importante que dans 100 µl de tampon ( $(6,9 \pm 1,4) \cdot 10^8$  CFU/ml pour 20 µl d'éluant,  $(4,6 \pm 1,6) \cdot 10^9$  CFU/ml pour 100 µl d'éluant). Le volume mort des billes étant de 80 µl, seule une élution avec un volume supérieur à 80 µl permet de mettre convenablement les billes à pH acide.

Les paramètres suivants ont donc été choisis pour le protocole final optimisé : incubation avec un volume de 80 µl de billes et élution par un volume de 100 µl de TFA 0,2 %.

## 6.7 Récapitulatif du protocole de couplage et d'immunocapture

Voici le protocole tel qu'il a été modifié avec toutes ces étapes d'optimisation et tel qu'il est appliqué pour toutes les expériences qui suivent.

### 6.7.1 Protocole de couplage des billes magnétiques

Le protocole pour le couplage des billes est le suivant. Pour 4 mg de billes, un volume de 40 µl de solution contenant les billes est prélevé. Un portoir magnétique (DynaMag<sup>TM</sup>-2, Invitrogen) adapté permet d'isoler contre une paroi du tube les billes magnétiques afin d'éliminer les différentes solutions sans perdre de billes. Les billes sont lavées à deux reprises avec 400 µl de tampon borate de sodium 0,1 M (pH=9,5). Après élimination du dernier surnageant les billes sont reprises dans 33 µl de tampon borate de sodium 0,1 M (pH=9,5). Un volume de 25 µl d'anticorps (FT100R) à 4 mg/ml dans du tampon borate de sodium est ajouté, suivi de 17 µl de sulfate d'ammonium 3 M, afin d'atteindre une concentration finale de 0,5 M. Le mélange est laissé à incuber une nuit à 37°C dans un Thermomixer® sous agitation. La réaction de couplage est bloquée par 400 µl d'éthanolamine à 50 mM dans du PBS Tween 0,05 %. Le mélange est incubé de nouveau une nuit à 37 °C sous agitation dans un Thermomixer®. Une fois le surnageant éliminé, les billes sont lavées à deux reprises par 400 µl de PBS Tween 0,05 %. Après élimination du dernier surnageant les billes sont reprises dans 80 µl de PBS Tween 0,05 % avant d'être stockées à 4°C avant utilisation.

### 6.7.2 Protocole d'immunocapture des bactéries actives

Le protocole conservé pour l'immunocapture est le suivant. Pour un volume de 1 ml d'échantillon, 4 mg de billes Dynabeads® MyOne<sup>TM</sup> Tosylactivated (Invitrogen) à 25 µg d'Ac/mg de billes soit 80 µl de billes sont utilisées. Les billes sont d'abord lavées à 3 reprises avec 400 µl de PBS stérile. Après avoir éliminé le dernier surnageant, l'échantillon est ajouté et laissé à incuber 1 h à température ambiante sous agitation sur un Thermomixer®. A la fin de l'incubation, le surnageant est éliminé et les billes sont lavées à deux reprises avec 400 µl de PBS stérile puis une fois avec 400 µl d'eau stérile. Après l'élimination du dernier surnageant, les bactéries sont éluées par 100 µl de TFA 0,2 %. Le pH de l'éluat est neutralisé par 1 µl d'hydroxyde d'ammonium 25 %.

## 7 Optimisation et validation de l'approche MS

Le MALDI-TOF a été choisi pour analyser les résultats de l'immunocapture. En effet, cet appareil permet de visualiser simplement et rapidement ce qui est présent dans un échantillon. L'échantillon et la matrice permettant la cristallisation sont déposés sur une cible MALDI. Le séchage prend quelques minutes à l'air libre et le tir MALDI quelques secondes. La lecture du résultat peut être directe - présence ou absence de spectre - ou peut nécessiter une analyse avec un logiciel pour une confirmation d'identification par exemple.

### 7.1 Analyse MALDI-TOF

#### 7.1.1 Matrice

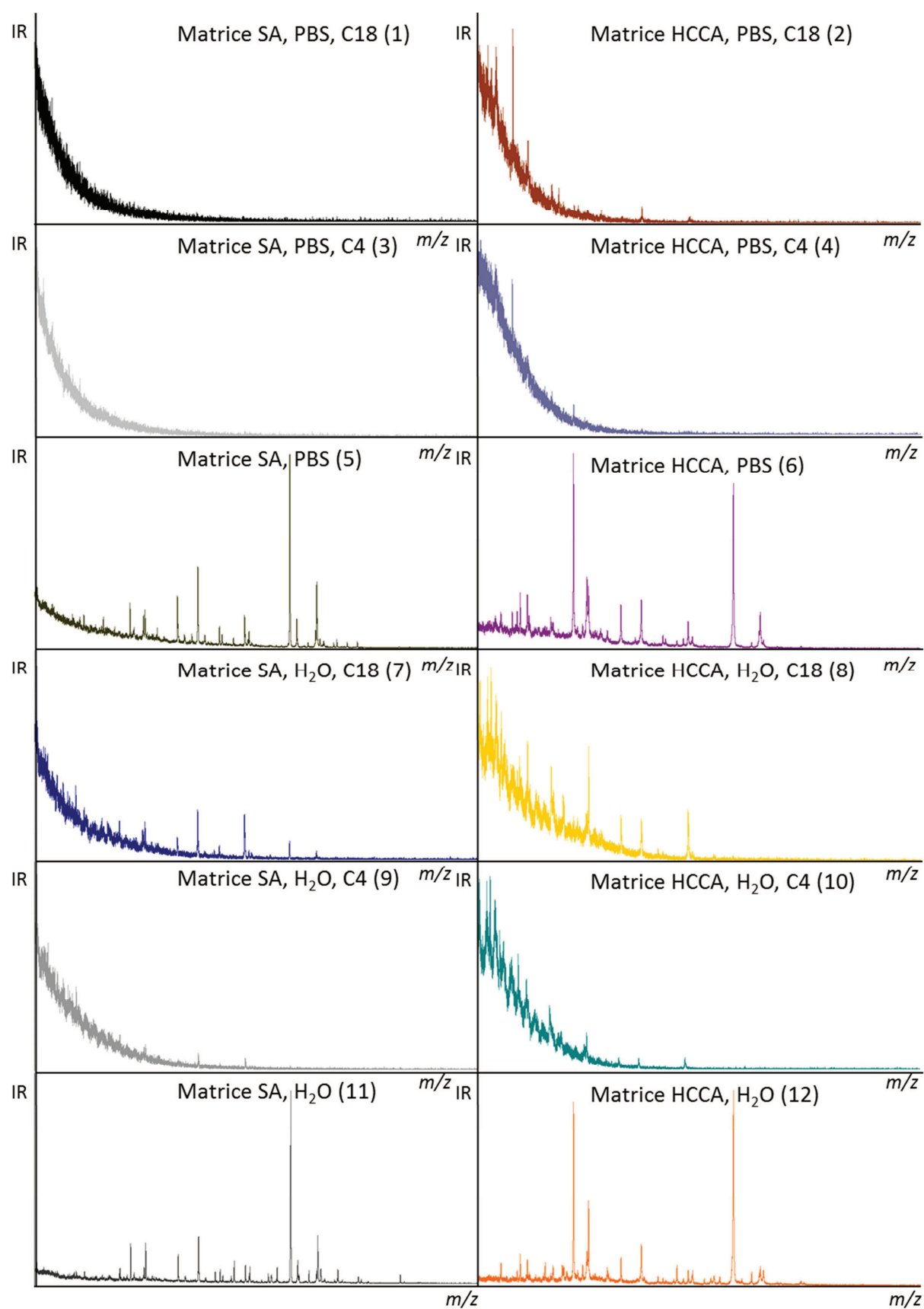
Pour la cristallisation des dépôts plusieurs types de matrices peuvent être utilisés. Nous avons choisi de travailler avec les matrices suivantes l'acide sinapinique (SA) et l'acide  $\alpha$ -cyano-4-hydroxycinnamique (HCCA). Les matrices se présentent sous forme d'une poudre que nous mettons à saturation dans 50 % ACN/1,25 % TFA pour la matrice SA et 50 % ACN/2,5 % TFA pour la matrice HCCA. Après agitation, seule la solution saturée est conservée pour les dépôts. Nous avons aussi travaillé avec des ZipTip® C<sub>4</sub> et C<sub>18</sub> (Millipore) pour voir si le signal pouvait être amélioré par une étape de dessalage. Les ZipTip® C<sub>4</sub> et C<sub>18</sub> sont des cônes qui s'adaptent aux pipettes et dont la pointe est garnie de résine de silice greffée en C<sub>4</sub> ou en C<sub>18</sub>. Les ZipTip® C<sub>4</sub> conviennent pour des protéines ou des peptides de masses comprises entre 3 et 100 kDa et les ZipTip® C<sub>18</sub> sont mieux adaptés pour des oligonucléotides, des protéines ou des peptides de masses inférieures à 50 kDa.

#### 7.1.2 Les différentes conditions

Dans le **Tableau 12** sont détaillées les différentes conditions utilisées pour chaque dépôt sur cible MALDI. Pour préparer les solutions de bactéries, nous avons réalisé une culture de *Francisella tularensis* subsp. *holarctica* LVS en milieu liquide. Après centrifugation de la culture séparée en deux tubes distincts, le culot bactérien est lavé à deux reprises avec du PBS stérile. Le culot bactérien est repris avec du PBS stérile ou de l'eau stérile de manière à obtenir une DO à 600 nm de 1 (soit une concentration d'environ 10<sup>9</sup> CFU/ml). Ensuite, les échantillons sont immédiatement inactivés à la chaleur avant d'être stockés à -80°C. L'ensemble des spectres acquis est rassemblé sur la **Figure 34**.

**Tableau 12** : Conditions de tampons, de ZipTip et de matrices utilisées pour faire les spectres MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS.

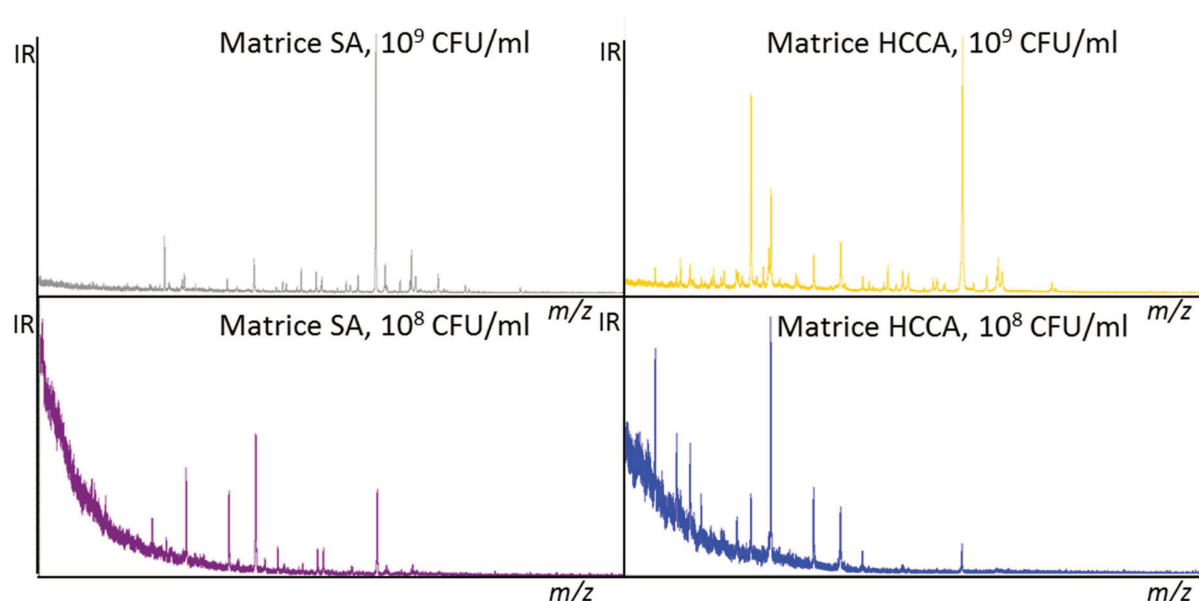
Tube	Bactéries en tampon :	ZipTip	Matrice
1	PBS	C18	SA
2	PBS	C18	HCCA
3	PBS	C4	SA
4	PBS	C4	HCCA
5	PBS	-	SA
6	PBS	-	HCCA
7	H <sub>2</sub> O	C18	SA
8	H <sub>2</sub> O	C18	HCCA
9	H <sub>2</sub> O	C4	SA
10	H <sub>2</sub> O	C4	HCCA
11	H <sub>2</sub> O	-	SA
12	H <sub>2</sub> O	-	HCCA



**Figure 34** : Spectres MALDI-TOF des bactéries selon les différentes conditions décrites dans le **Tableau 12**, présentés dans la gamme de  $m/z$  de 2 000 à 15 000 en fonction de leur intensité relative (IR).

De nouveau, nous pouvons constater une nette diminution du bruit de fond lorsque les bactéries sont dans de l'eau. En effet, le spectre de *Francisella tularensis* subsp. *holarctica* LVS dans de l'eau MilliQ avec la matrice SA est mieux résolu que le même échantillon dans du PBS. Nous pouvons également constater que les intensités des différents pics varient selon la matrice utilisée, avec des pics plus intenses sur la gamme de  $m/z$  de 4 000 à 6 000 pour la matrice HCCA. Les ZipTip ne permettent pas d'observer plus de protéines ou d'obtenir une meilleure sensibilité : les spectres sont vierges.

Des dilutions des échantillons au dixième dans de l'eau ont été déposées avec les matrices SA et HCCA pour déterminer si la limite de sensibilité est améliorée avec l'une ou l'autre des matrices. Les résultats obtenus sont présentés sur la **Figure 35**.



**Figure 35** : Spectres MALDI-TOF des bactéries à deux concentrations dans de l'eau avec les matrices SA et HCCA, présentés dans la gamme de  $m/z$  de 2 000 à 15 000 en fonction de leur intensité relative (IR).

Il est normal d'observer des différences entre les profils pour deux matrices différentes puisque l'ionisation des protéines diffère selon les caractéristiques des matrices utilisées pour la cristallisation. Bruker recommande d'utiliser la matrice HCCA comme matrice pour le Biotyper, il en est de même pour le Vitek<sup>®</sup> MS de bioMérieux. En effet, pour un logiciel qui se base sur la reconnaissance des profils MALDI-TOF pour identifier les microorganismes, il est indispensable d'utiliser toujours la même matrice afin d'avoir les profils les plus reproductibles possibles. Lorsque nous observons les spectres acquis, nous constatons qu'il y a plus de pics de faible  $m/z$  avec la matrice HCCA et inversement plus de pics de fort rapports  $m/z$  avec la matrice SA. Les dilutions de bactéries déposées avec les deux matrices montrent qu'il n'y a pas de différences significatives pour la limite de détection des bactéries. L'élimination du maximum de sels lors de la reprise des échantillons est indispensable pour diminuer le bruit de fond.



### 7.1.3 Calibration

Nous avons utilisé deux types de calibrants : Protein Calibration Standard I (PCSI, Bruker) et Bruker Bacterial Test Standard (BTS, Bruker). Le premier calibrant, PCSI, est constitué de quatre protéines visibles sous leur forme monochargée et dont deux sont visibles également sous leur forme doublement chargée, l'insuline ( $m/z$  : 5 734,51), l'ubiquitine I ( $m/z$  : 8 565,76), le cytochrome C ( $m/z$  : 12 360,97 et 6 180,99) et la myoglobine ( $m/z$  : 16 952,30 et 8 476,65). Le BTS est constitué d'un extrait de *Escherichia coli* DH5 alpha auquel est ajouté deux autres protéines, la RNaseA et la myoglobine, de manière à couvrir un large spectre de masses pour la calibration. Les protéines prises en compte pour *E. coli* sont des protéines ribosomales sous leur forme monochargée ou doublement chargée (RL29  $[M+2H]^{2+}$ ,  $m/z$  : 3 637,8 ; RS32  $[M+H]^+$ ,  $m/z$  : 5 096,8 ; RS34  $[M+H]^+$ ,  $m/z$  : 5 381,4 ; RS33meth  $[M+H]^+$ ,  $m/z$  : 6 255,4 ; RL29  $[M+H]^+$ ,  $m/z$  : 7 274,5 ; RS19  $[M+H]^+$ ,  $m/z$  : 10 300,1 Da RNaseA  $[M+H]^+$ ,  $m/z$  : 13 683,2 ; Myoglobine  $[M+H]^+$ ,  $m/z$  : 16 952,3).

Nous utilisons le calibrant PCSI de manière courante au laboratoire. Il est possible de le mélanger à un échantillon pour en faire la calibration interne. Le calibrant BTS, est celui qui est recommandé par Bruker pour les analyses avec le Biotyper. Comme nous souhaitons faire de la comparaison de profils protéiques, nous avons décidé d'utiliser le calibrant BTS pour nous rapprocher le plus possible du protocole pour le Biotyper.

## 7.2 Traitement des données MALDI-TOF

Après avoir déterminé les biomarqueurs permettant de différencier les différentes espèces et sous-espèces de *Francisella* (cf. 1<sup>ère</sup> partie : Identification de *Francisella tularensis* jusqu'au niveau sous-espèce par la détection de biomarqueurs spécifiques), nous avons pensé utiliser cette technique pour identifier les bactéries après enrichissement. Après plusieurs expériences nous nous sommes aperçus que parfois un spectre était présent et confirmait la présence d'une bactérie, mais que les 3 biomarqueurs sélectionnés n'étaient pas toujours visibles et donc ne pouvait valider la présence de la bactérie dans l'échantillon. Le laboratoire ne possédant pas le logiciel Biotyper d'identification de microorganisme, il a fallu mettre au point un protocole pour valider la présence des bactéries après enrichissement.

### 7.2.1 Traitement des spectres bruts

Pour traiter les spectres bruts obtenus avec le MALDI-TOF (Biflex IV, Bruker) contrôlé par le logiciel FlexControl (Bruker), nous avons utilisé le logiciel FlexAnalysis (Bruker). FlexControl permet de fixer les paramètres d'acquisition du MALDI-TOF tels que le mode de détection (linéaire ou réflectron), la puissance du laser ou la gamme de masses analysée. Ici nous nous sommes placés en mode linéaire, plus adapté pour l'analyse de protéines entières, sur la gamme de  $m/z$  entre 2 000 et 20 000.

FlexAnalysis permet de retraiter les spectres en soustrayant la ligne de base et en lissant les spectres. Il est également possible de recalibrer les spectres en utilisant la calibration interne et d'extraire les listes des masses selon 4 modes différents de traitement des spectres. Le mode « Apex » utilise la première et la seconde dérivée pour déterminer la valeur maximum

du pic. Le mode « Sum » est un mode de détection de pic rapide, il utilise des pseudos dérivées, si la pente dépasse un certain seuil le logiciel considère qu'il y a un pic. Le mode « SNAP » identifie des motifs isotopiques et calcule les masses monoisotopiques de manière exacte ce qui peut être intéressant pour le séquençage *de novo*. Le mode « Centroid » se base sur le même principe que le mode Apex, mais utilise un certain pourcentage de la hauteur du pic pour calculer la valeur maximum du pic. Ce pourcentage est désigné par le paramètre « centroid percent high ». C'est ce dernier mode que nous avons choisi. Nous faisons l'extraction des 200 pics les plus intenses dont le signal sur bruit (S/N) est supérieur à 6 avec un seuil de 80 %. Ce processus d'extraction et de sauvegarde des listes de masses a été automatisé. En effet, sans cette automatisation chaque spectre doit être traité individuellement. Le script de traitement de spectres programmé en BASIC est consultable en **Annexe 3**.

### 7.2.2 Algorithme de comparaison des profils protéiques

Les biomarqueurs ne peuvent pas être le seul moyen de confirmer la présence de la bactérie. En effet, à la limite de sensibilité du MALDI-TOF, seuls quelques pics peuvent apparaître ce qui montre que la bactérie est présente sans pour autant que les pics des biomarqueurs soient présents.

Nous avons envisagé d'utiliser une analyse en composantes principales (ACP) pour comparer les profils de protéines. Cette analyse consiste à calculer de nouvelles variables permettant d'observer les individus de manière plus claire sur une représentation bi ou tridimensionnelle. Pour cela les variables sont d'abord centrées et réduites, c'est-à-dire que leur moyenne est nulle et que leur variance est égale à 1. Ainsi, les données sont indépendantes de l'unité ou de l'échelle choisie et les variables ont les mêmes moyennes et les mêmes dispersions. De nouveaux axes sont créés à partir de combinaisons linéaires des anciennes variables : ce sont les axes factoriels. Il y aura autant de nouveaux axes qu'il y a d'anciennes variables, tous les axes n'auront pas la même importance. Le premier axe, ou axe d'inertie est celui qui représente le plus grand allongement du nuage statistique. La projection des points sur cet axe est la plus dispersée, c'est sur cet axe que l'on verra les plus grandes différences entre les individus et donc également les similarités. Les axes suivants représentent les autres directions d'allongement du nuage. Pour choisir les axes factoriels il faut s'attacher au calcul du pourcentage d'inertie. Plus ce pourcentage est élevé plus l'axe est pertinent. Généralement, on choisit le nombre d'axes nécessaire pour atteindre un seuil de pourcentage (80 % par exemple), ce qui correspond la plupart du temps aux deux premiers axes. Le calcul et la représentation de l'ACP nécessitent des logiciels spécialisés tels que MATLAB® (payant), Octave (gratuit), R (gratuit) ou XLSTAT, un complément d'Excel (payant). C'est ce dernier que nous avons choisi d'utiliser pour faire les ACP.

L'ACP permet de bien représenter les différences entre les échantillons malgré le grand nombre de variables qui les caractérisent. De plus, les pourcentages d'inertie fournissent un indice de confiance sur le choix des axes factoriels et donc sur la pertinence de la projection des données sur les axes de la représentation graphique. En termes de choix de variables

pour utiliser ces analyses en composante principale, la sélection des pics compris dans un intervalle de 5  $m/z$  autour de pics définis comme des biomarqueurs de la souche étudiée permet de bien observer les différences entre les échantillons de deux souches distinctes. Même si cette méthode permet de différencier une souche de *Francisella* d'une souche de *E. coli*, elle ne permet pas d'évaluer la qualité des spectres d'une même souche les uns par rapport aux autres.

Pour avoir une confirmation objective de la présence de la bactérie nous avons mis au point un algorithme de comparaison des profils permettant d'attribuer un score de probabilité de présence de *Francisella* aux spectres obtenus après l'enrichissement. Pour cela, nous nous sommes basés sur la publication de l'équipe de Jarman (Jarman et al., 2000). Dans cette publication est décrite une méthode permettant d'automatiser l'identification de microorganisme à partir de spectres MALDI-TOF. Le principe est le suivant : pour un organisme choisi, un certain nombre de spectres sont acquis dans la gamme de  $m/z$  allant de 2 000 à 20 000 et les pics présents parmi les 200 pics les plus intenses dans 70 % des spectres vont constituer les pics de référence de l'organisme. A chaque pic de référence est attribuée une probabilité de présence ( $p_i$ ) qui dépend de la fréquence d'apparition de ce pic dans les spectres qui ont permis de constituer le spectre de référence. Nous avons procédé à la création d'une liste de pics de référence pour *Francisella tularensis* subsp. *holarctica* LVS à partir de 19 spectres. Ce sont 46 valeurs de  $m/z$  qui ont été retenues avec des rapports  $m/z$  compris entre 2 590 et 10 365. La valeur des rapports  $m/z$  retenus est calculée comme la moyenne de tous les rapports  $m/z$  d'un pic donné. Une erreur sur les valeurs des  $m/z$  est prise en compte, nous la faisons varier entre 300 et 500 ppm. Les spectres dont nous voulons faire l'identification sont également acquis dans la gamme de  $m/z$  variant de 2 000 à 20 000 et les 200 pics les plus intenses sont extraits à l'aide de FlexAnalysis. Pour chaque pic de référence un intervalle de masse est calculé avec l'erreur. Pour chaque intervalle de référence, l'algorithme va vérifier s'il y a un pic présent dans cet intervalle dans le spectre à comparer. Un vecteur de présence et un vecteur d'absence de pics dans les intervalles de références vont être créés. Ils seront multipliés par le vecteur des probabilités de présence. Le score final est calculé par l'

**Équation 3** comme une combinaison linéaire de la probabilité que tous les pics de référence soient observés et de la probabilité qu'aucun pic ne soit observé.

**Équation 3** : Calcul du score de probabilité de présence ( $da$ ). Tirée de (Jarman et al., 2000).

$$da(k) = 1 - \prod_{i \in M^c} p_i \left[ 1 - \prod_{i \in M} (1 - p_i) \right]$$

$p_i$  : probabilité de présence du pic  $i$  basée sur la fréquence d'apparition du pic  $i$  dans les spectres qui ont permis de constituer le spectre de référence.

$M$  : ensemble des pics de référence absents dans le spectre de l'échantillon inconnu.

$M^C$  : ensemble des pics de référence présents dans le spectre de l'échantillon inconnu.

Les scores (da) obtenus sont compris entre 0 et 1. Le score exprime la probabilité de présence de *Francisella tularensis* subsp. *holarctica* LVS dans l'échantillon. La signification des scores est décrite dans le **Tableau 13**.

**Tableau 13** : Signification des scores de l'algorithme d'identification. Tiré de (Jarman et al., 2000)

Score (da)	Signification
< 0,01	Très improbable
$0,01 \leq da < 0,05$	Improbable
$0,05 \leq da < 0,15$	Impossible de conclure
$0,15 \leq da < 0,7$	Probable
$\geq 0,7$	Très probable

Nous avons validé cet algorithme sur des spectres MALDI-TOF de plusieurs types de bactéries différentes de *Francisella tularensis* subsp. *holarctica* LVS dont les genres ont été déterminés par séquençage de l'ARN 16S. Les scores présents dans le **Tableau 14** ont été calculés par l'algorithme de reconnaissance de *Francisella tularensis* subsp. *holarctica* LVS sur 3 spectres MALDI-TOF de chaque genre de bactéries qui ont été acquis au laboratoire. L'ensemble des scores obtenus était nul, ce qui valide l'approche de reconnaissance de *Francisella tularensis* subsp. *holarctica* LVS avec cet algorithme.

**Tableau 14** : Résultats de l'algorithme de reconnaissance de *Francisella tularensis* subsp. *holarctica* LVS sur des bactéries de genres différents de *Francisella*.

Genre	Score
<i>Zobelia</i>	0
<i>Lutibacterium</i>	0
<i>Staphylococcus</i>	0
<i>Paracoccus</i>	0
<i>Ruegeria</i>	0
<i>Acinetobacter</i>	0
<i>Vibrio</i>	0
<i>Shewanella</i>	0
<i>Pseudoalteromonas</i>	0
<i>Mycobacterium</i>	0
<i>Vibrio</i>	0
<i>Muricauda</i>	0
<i>Streptomyces</i>	0
<i>Aquimarina</i>	0

Pour coder cet algorithme nous avons utilisé Octave (<http://www.gnu.org/software/octave/>), un langage de programmation en libre accès très proche de Matlab qui est lui payant. Le script est présenté en **Annexe 4**.

## 8 Banc d'essai

Nous allons maintenant développer les résultats obtenus sur différents types d'échantillons. Les résultats sont présentés sous forme d'histogramme. Pour chaque série d'échantillons le score et le nombre signaux  $m/z$  correspondant au spectre type de la bactérie sont donnés. Le seuil d'identification pour la bactérie est un score de 0,15 et le score est calculé lorsqu'au moins 9 pics du spectre de référence sont trouvés, soit 20 % du total des pics de référence. Ceci permet d'éviter que la présence de quelques pics, sur le spectre d'un échantillon inconnu, et dont la probabilité de présence est élevée dans le spectre de référence, suffise pour faire une identification.

### 8.1 Test de matrices

Nous avons testé le protocole d'immunocapture sur plusieurs types de matrices : du PBS, de l'eau minérale (Evian), du lait UHT demi-écrémé et des prélèvements d'air.

#### 8.1.1 Essais en matrice PBS

Le **Tableau 15** représente les différentes conditions d'immunocapture qui ont été menées en prenant comme matrice le PBS. Trois analyses totalement indépendantes sont réalisées sur trois jours différents. Ces trois répliquats d'immunocapture ont été réalisés à trois concentrations différentes.

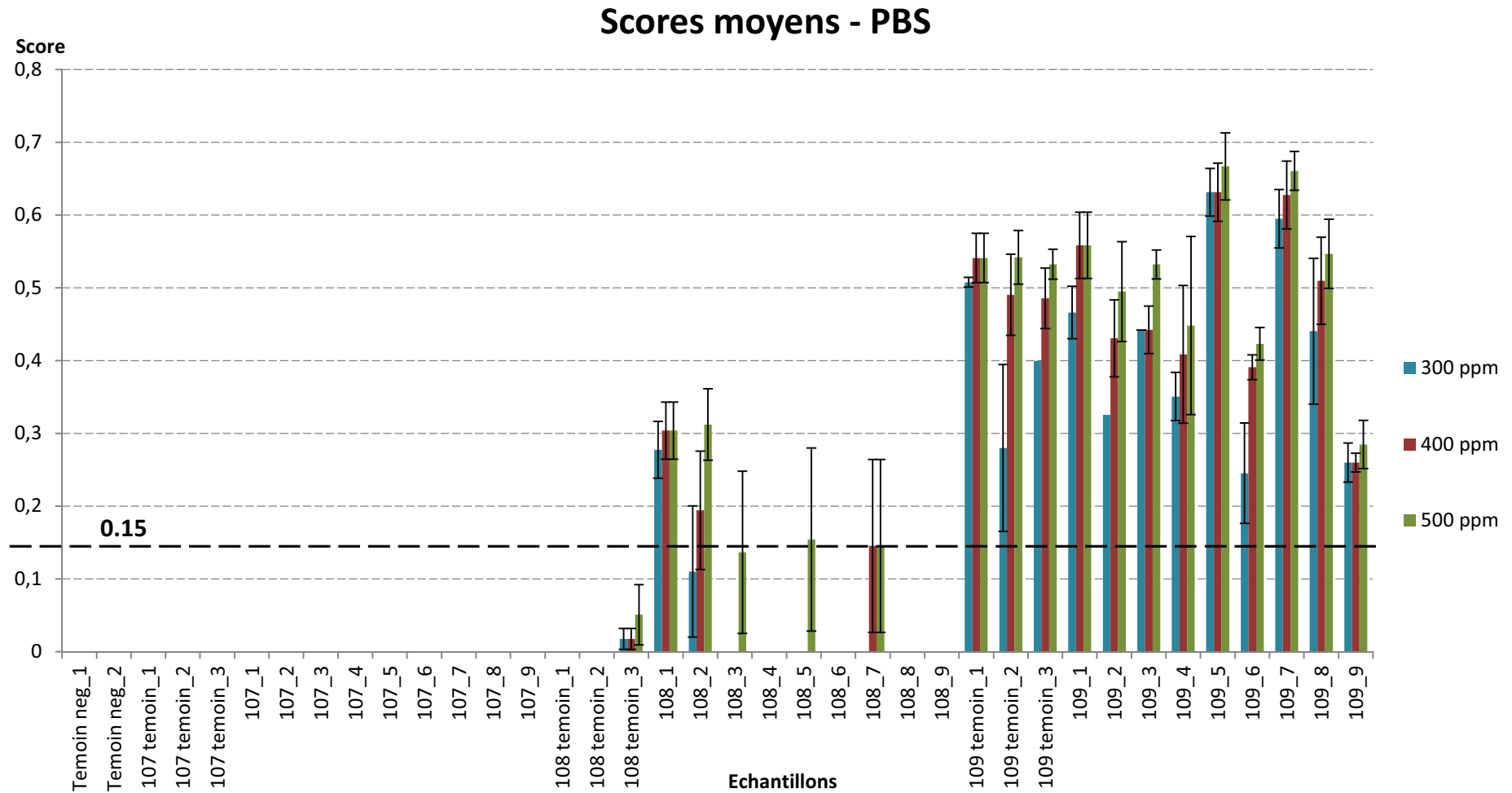
**Tableau 15 :** Récapitulatif des échantillons préparés en PBS. Tous les échantillons subissent une immunocapture à l'exception des échantillons nommées « Témoin ».

Concentration	10 <sup>9</sup> CFU/ml	10 <sup>8</sup> CFU/ml	10 <sup>7</sup> CFU/ml
Jour 1	109_1	108_1	107_1
	109_2	108_2	107_2
	109_3	108_3	107_3
<b>Témoin</b>	109Témoin_1	108Témoin_1	107Témoin_1
Jour 2	109_4	108_4	107_4
	109_5	108_5	107_5
	109_6	108_6	107_6
<b>Témoin</b>	109Témoin_2	108Témoin_2	107Témoin_2
Jour 3	109_7	108_7	107_7
	109_8	108_8	107_8
	109_9	108_9	107_9
<b>Témoin</b>	109Témoin_3	108Témoin_3	107Témoin_3

Les témoins sont constitués de solutions de bactéries aux trois concentrations et sont directement inactivés sans immunocapture. Les éluats obtenus après l'immunocapture et les témoins sont déposés sur la cible MALDI avec les paramètres suivants : 1 µl d'éluat déposé puis 1 µl de matrice HCCA déposé. Les tirs MALDI ont été réalisés sur la gamme de  $m/z$  de 2 000 à 20 000 et les spectres sont la somme de 6 salves de 40 tirs. Pour chaque éluat trois spectres sont acquis (répliquats analytiques). La moyenne des trois scores obtenus pour chaque échantillon est représentée sur la **Figure 36**. Les scores sont également présentés

pour trois types de tolérance : 300, 400, et 500 ppm. Cette gamme d'erreur possible est appliquée sur la fenêtre de recherche des pics de référence sur le spectre de l'échantillon inconnu. En effet, les spectres n'ont pas de calibration interne mais seulement une calibration externe ce qui peut entraîner un décalage de l'ensemble des masses et donc éventuellement des erreurs d'identification.

Les échantillons avec une concentration initiale de  $10^9$  CFU/ml sont identifiables avant l'immunocapture (109Temoin\_1, 109Temoin\_2, 109Temoin\_3), ce qui correspond à ce que nous avons observé précédemment. En effet, à  $10^9$  CFU/ml nous sommes proches de la limite de détection du MALDI, ce qui explique que le témoin soit identifiable. Les 9 résultats d'immunocapture mènent à une identification avec une moyenne de score de  $0,51 \pm 0,15$ . Pour les échantillons avec une concentration initiale de  $10^8$  CFU/ml (108Temoin\_1, 108Temoin\_2, 108Temoin\_3) l'identification n'est pas possible, le score pour un échantillon est inférieur à 0,1 et pour les deux autres échantillons le nombre de pics sélectionné est inférieur à 9 donc aucun score n'est donné. Ceci s'explique par le fait que nous sommes en dessous de la limite de détection du MALDI-TOF pour ce type d'échantillons. Les résultats de l'immunocapture pour cette concentration ( $10^8$  CFU/ml) sont peu reproductibles avec 4 échantillons sur 9 dont les spectres sont de qualité suffisante avec une erreur de 500 ppm pour valider l'identification. Ceci peut s'expliquer par le fait qu'il existe une variabilité au moment de la préparation des échantillons sur la concentration réelle de bactéries (mesure de DO à 600 nm). De plus, la concentration attendue après l'immunocapture est également à la limite de détection. A  $10^7$  CFU/ml, aucun spectre ne permet l'identification des bactéries, ce qui apparaît correct au vu des limites de détection préalablement établies. Nous pouvons également constater que lorsque l'erreur sur la masse passe de 300 à 500 ppm le score d'identification est souvent amélioré. En effet, une fois déposés sur la cible MALDI les échantillons ne bénéficient que d'une calibration externe. Nous avons observé que les dépôts ne sont pas toujours homogènes ce qui entraîne des variations dans les masses pouvant aller jusqu'à plusieurs unités de  $m/z$ .



**Figure 36** : Scores moyens sur 3 tirs sur un dépôt d'éluat. Les scores supérieurs à 0,15 valident l'identification. Erreurs : en bleu, 300 ppm ; en rouge, 400 ppm ; en vert, 500 ppm.

### 8.1.2 Essais en matrice lait UHT demi-écrémé

Les échantillons qui ont été traités dans du lait sont détaillés dans le **Tableau 16**. Pour préparer les échantillons, nous avons procédé de la manière suivante : pour obtenir la concentration voulue dans 1 ml de lait, nous avons préparé des tubes contenant la quantité de bactéries correspondant à la quantité de bactéries devant être obtenue dans l'échantillon de lait. Ces volumes de solution bactérienne ont été soit centrifugés pour ne garder que le culot de bactéries et le resuspendre dans 1 ml de lait (volume de 100 µl et 1 ml), soit directement ajouté au volume de lait, le volume de solution bactérienne étant négligeable (10 µl).

**Tableau 16** : Méthode de préparation des échantillons de lait.

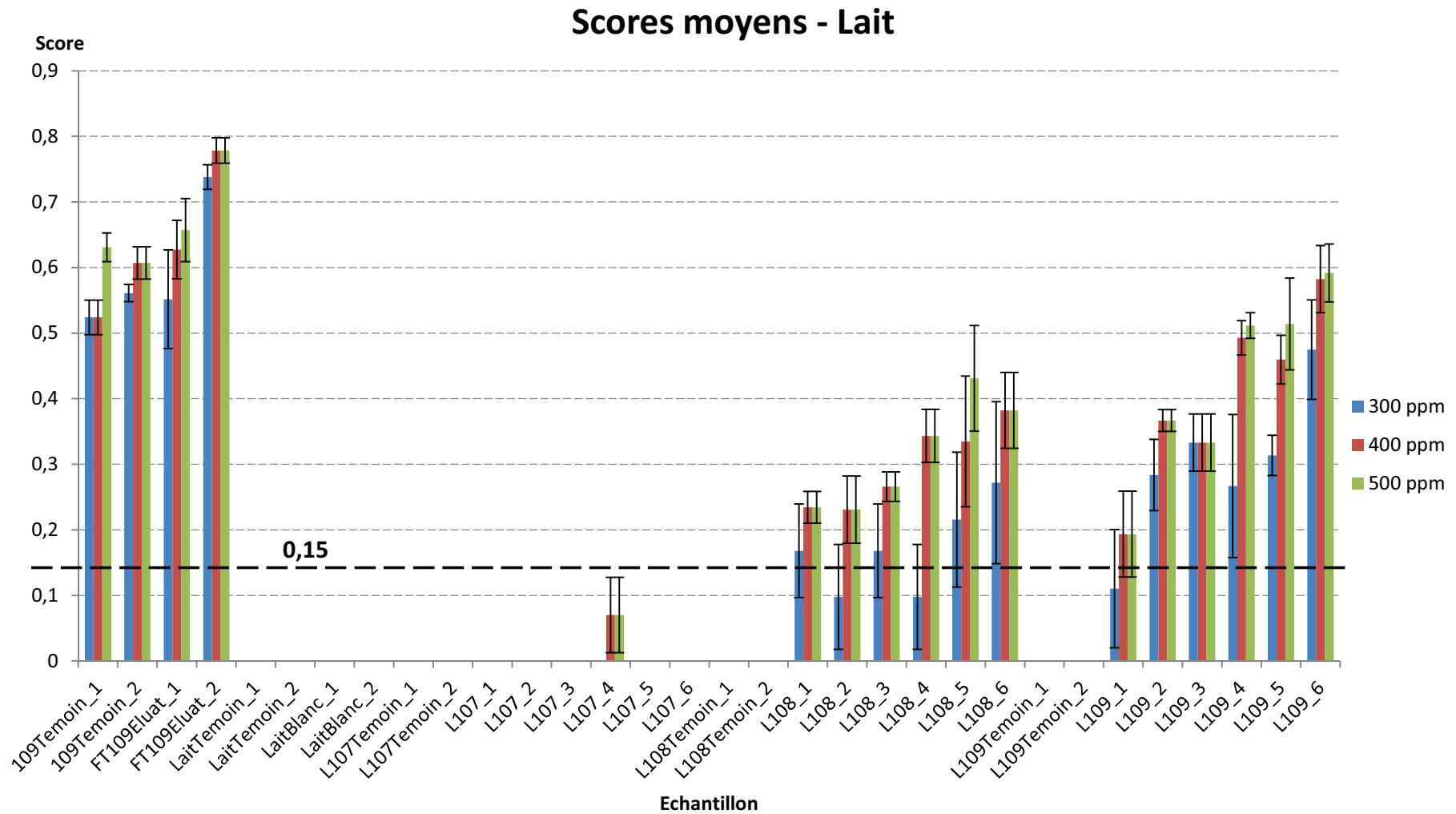
Jour 1	Jour 2	Volume lait UHT	Volume H <sub>2</sub> O MilliQ	Culot de X ml de FT à 10 <sup>9</sup> CFU/ml	Concentration avant immunocapture
L109_1 L109_2 L109_3	L109_4 L109_5 L109_6	900 µl	100 µl	1 ml à culotter	10 <sup>9</sup> CFU/ml
L108_1 L108_2 L108_3	L108_4 L108_5 L108_6	900 µl	100 µl	100 µl à culotter	10 <sup>8</sup> CFU/ml
L107_1 L107_2 L107_3	L107_4 L107_5 L107_6	900 µl	100 µl	10 µl	10 <sup>7</sup> CFU/ml

La **Figure 37** représente les scores moyens de trois spectres pour chaque échantillon de l'immunocapture dans du lait. Sur la gauche de l'histogramme sont présents les différents témoins dont la composition est détaillée dans le **Tableau 17**.

**Tableau 17** : Composition des différents témoins d'immunocapture.

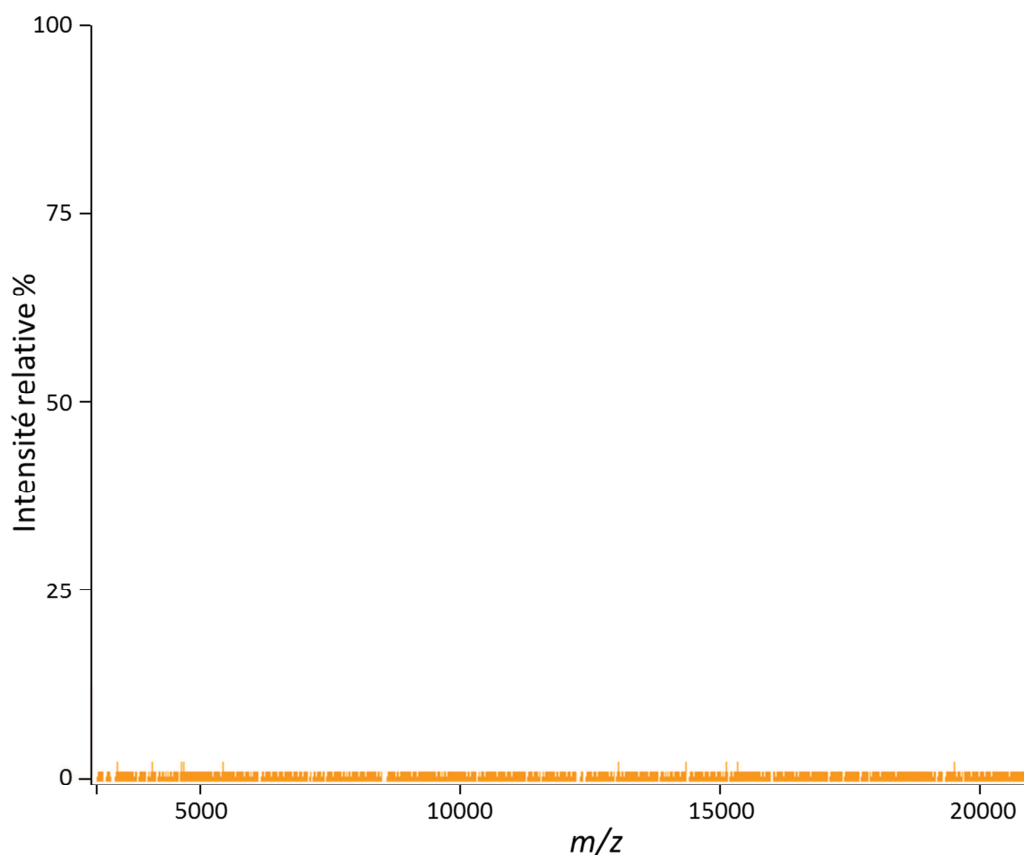
Tubes	Volume lait UHT	Volume H <sub>2</sub> O MilliQ	Volume de FT à 10 <sup>9</sup> CFU/ml	Concentration avant immunocapture	Immunocapture
109_Temoin	-	-	1 ml	10 <sup>9</sup> CFU/ml	Non
FT109_Eluat	-	-	1 ml	10 <sup>9</sup> CFU/ml	Oui
Lait_Temoin	900 µl	100 µl	-	-	Oui
Lait_Blanc	900 µl	100 µl	-	-	Non
L107_Temoin	900 µl	100 µl	10 µl	10 <sup>7</sup> CFU/ml	Non
L108_Temoin	900 µl	100 µl	100 µl (à culotter)	10 <sup>8</sup> CFU/ml	Non
L109_Temoin	900 µl	100 µl	1 ml (à culotter)	10 <sup>9</sup> CFU/ml	Non





**Figure 37** : Scores moyens des immunocapture dans du lait. L'identification est validée lorsque le score est supérieur à 0,15. Erreurs : en bleu, 300 ppm ; en rouge, 400 ppm ; en vert, 500 ppm.

Les témoins permettent de confirmer (i) que les bactéries sont bien présentes dans l'échantillon et que l'immunocapture a fonctionné (FT109\_Eluat), (ii) que le lait seul ou le lait ayant subi une immunocapture n'ont pas un signal au MALDI-TOF qui pourrait être confondu avec celui de la bactérie, (iii) que l'immunocapture apporte une amélioration au signal par rapport aux échantillons de bactéries dans le lait analysées sans immunocapture. La **Figure 38** montre le spectre d'un échantillon témoin, contenant du lait et la bactérie à  $10^9$  CFU/ml (L109Témoin). La **Figure 37** montre qu'aucune des trois solutions témoins n'obtient un score suffisant ( $\geq 0,15$ ) pour une identification. En effet, le lait contient un fort taux de caséine qui masque le signal produit par les protéines des bactéries présentes dans le lait, même à  $10^9$  CFU/ml. Il est donc nécessaire d'enrichir les bactéries par rapport à la caséine du lait. Après immunocapture à  $10^8$  CFU/ml et  $10^9$  CFU/ml, les éluats montrent des identifications pour les 6 échantillons avec des scores moyens de  $0,42 \pm 0,16$  et  $0,31 \pm 0,12$  respectivement. A  $10^7$  CFU/ml, il n'y a pas d'identification après l'immunocapture, ce qui est normal car la concentration attendue est inférieure à la limite de détection. Ici l'immunocapture se montre bénéfique pour s'affranchir de la caséine présente dans le lait et l'immunocapture fonctionne correctement malgré la présence de caséine.



**Figure 38** : Spectre MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS à  $10^9$  CFU/ml dans du lait U.H.T demi-écrémé.

### 8.1.3 Essais en matrice eau d'Evian

Le **Tableau 18** représente la manière dont sont préparés les échantillons pour les essais dans l'eau d'Evian. Pour préparer les échantillons nous avons procédé de la manière suivante : pour obtenir la concentration voulue dans 1 ml d'eau d'Evian, nous avons préparé des tubes contenant la quantité de bactéries correspondant à la quantité de bactéries devant être obtenue dans l'échantillon d'eau d'Evian. Ces volumes de solution bactérienne ont soit été centrifugés pour ne garder que le culot de bactéries et le resuspendre dans 1 ml d'eau d'Evian (volume de 100 µl et 1 ml), soit directement ajouté au volume d'eau d'Evian, le volume de solution bactérienne étant négligeable (10 µl). Nous avons également ajouté du chlorure de sodium pour obtenir une concentration finale en sels de 150 mM afin d'être dans des conditions similaires aux immunocaptures dans du PBS. En effet, lors d'études préliminaires nous nous sommes aperçus que de meilleurs résultats d'immunocapture étaient obtenus en ajoutant ce chlorure de sodium.

**Tableau 18** : Méthode de préparation des échantillons d'eau d'Evian.

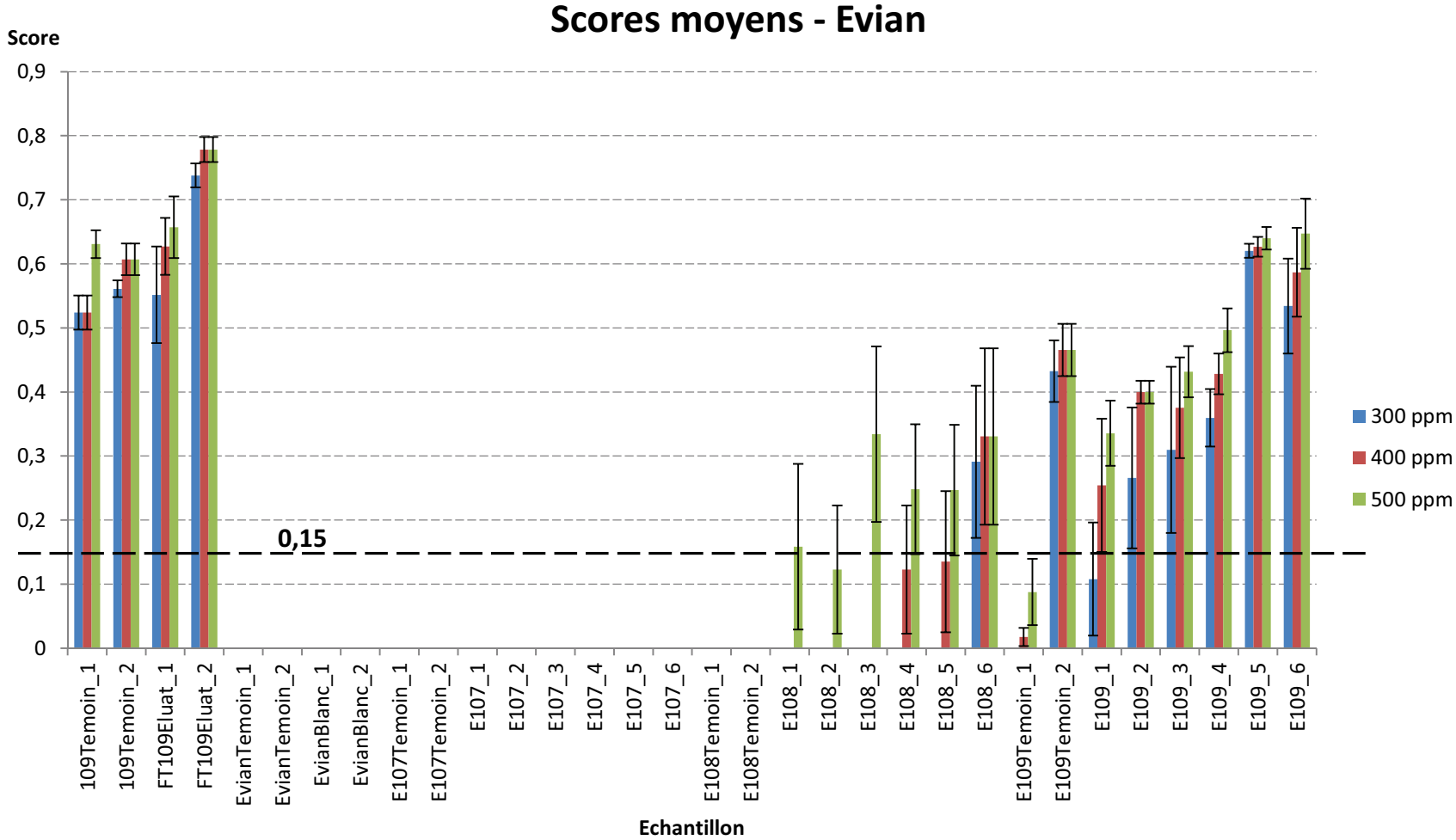
Jour 1	Jour 2	Volume eau Evian	Volume 1.5 M NaCl	Volume de FT à 10 <sup>9</sup> CFU/ml	Concentration avant immunocapture
E109-1	E109-4	900 µl	100 µl	1 ml à culotter	10 <sup>9</sup> CFU/ml
E109-2	E109-5				
E109-3	E109-6				
E108-1	E108-4	900 µl	100 µl	100 µl à culotter	10 <sup>8</sup> CFU/ml
E108-2	E108-5				
E108-3	E108-6				
E107-1	E107-4	900 µl	100 µl	10 µl	10 <sup>7</sup> CFU/ml
E107-2	E107-5				
E107-3	E107-6				

Le **Tableau 19** détaille la composition des différents témoins d'immunocapture utilisés dans cet essai.

**Tableau 19** : Composition des différents témoins d'immunocapture.

Tubes	Volume Evian	Volume 1.5 M NaCl	Volume de FT à 10 <sup>9</sup> CFU/ml	Concentration avant immunocapture	Immunocapture
109_Temoi	-	-	1 ml	10 <sup>9</sup> CFU/ml	Non
FT109_Eluat	-	-	1 ml	10 <sup>9</sup> CFU/ml	Oui
Evian_Temoi	900 µl	100 µl	-	-	Oui
Evian_Blanc	900 µl	100 µl	-	-	Non
E107_Temoi	900 µl	100 µl	10 µl	10 <sup>7</sup> CFU/ml	Non
E108_Temoi	900 µl	100 µl	100 µl (à culotter)	10 <sup>8</sup> CFU/ml	Non
E109_Temoi	900 µl	100 µl	1 ml (à culotter)	10 <sup>9</sup> CFU/ml	Non

Les résultats de cet essai sont représentés sur la **Figure 39**. Les témoins montrent que l'immunocapture fonctionne et que l'eau d'Evian seule et après une immunocapture ne présente pas de signal comparable à celui de la bactérie. A  $10^9$  CFU/ml un seul des deux témoins est identifié, cela est dû soit à un problème de calibration (les signaux  $m/z$  sont présents mais hors des fenêtres de recherche de l'algorithme) soit à une différence de concentration faisant que le témoin avant l'immunocapture est sous le seuil de détection du MALDI-TOF. Les résultats de l'immunocapture sont tous identifiés à cette concentration, avec un score moyen de  $0,49 \pm 0,14$ . A  $10^8$  CFU/ml, le témoin n'est pas identifié, et 5 échantillons sur 6 sont identifiés lorsque la tolérance est fixée à 500 ppm avec un score moyen de  $0,43 \pm 0,07$ .



**Figure 39** : Scores moyens des immunocapture dans de l’Evian. L’identification est validée lorsque le score est supérieur à 0,15. Erreurs : en bleu, 300 ppm ; en rouge, 400 ppm ; en vert, 500 ppm.

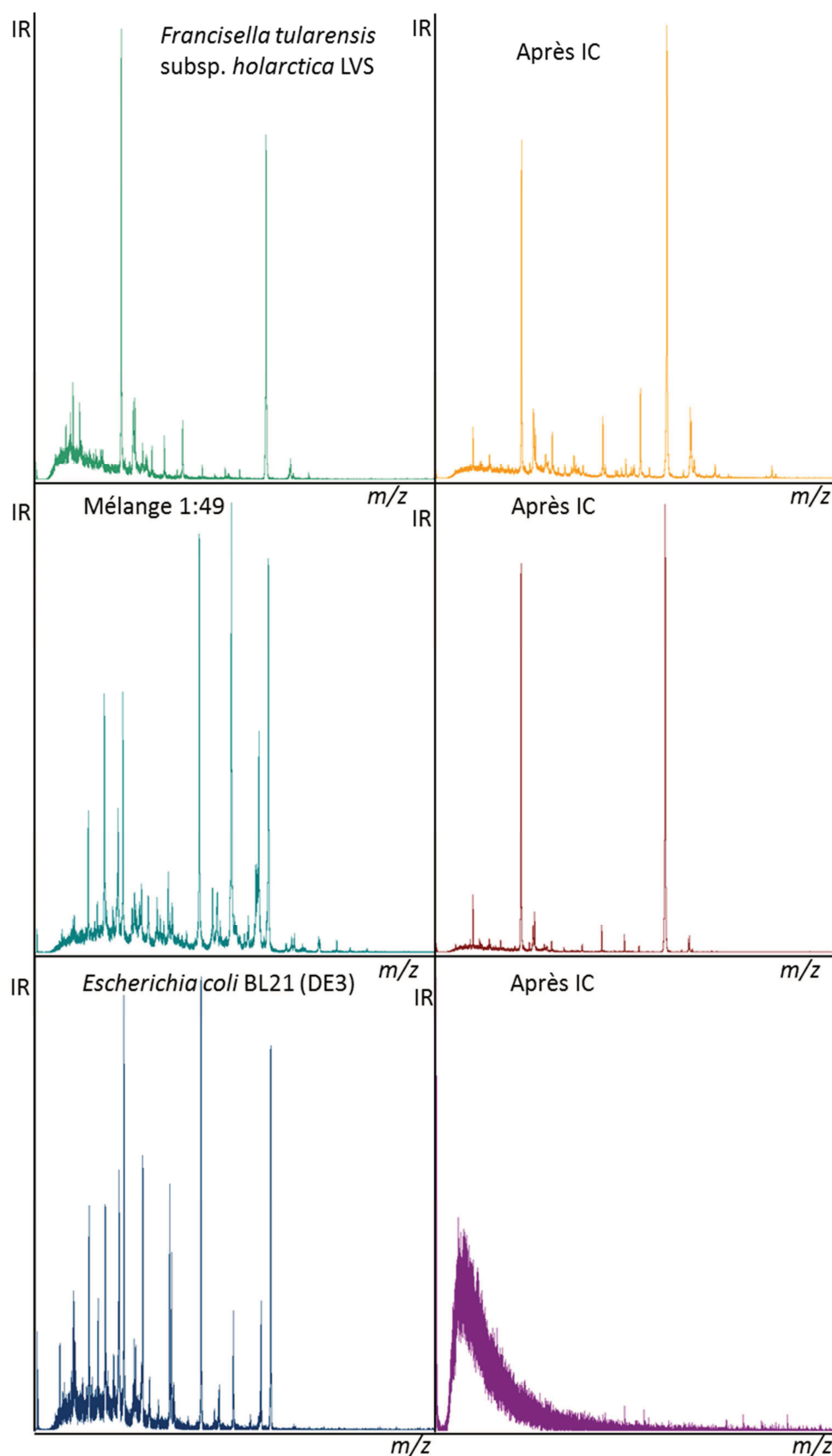
## 8.2 Mélanges hétérogènes de bactéries

Nous avons ensuite produit des mélanges hétérogènes de bactéries dans du PBS où les *Francisella tularensis* subsp. *holarctica* LVS sont en quantité inférieure à *Escherichia coli* BL21 (DE3). Cela permet de simuler le cas d'un échantillon environnemental où notre cible serait en très petite quantité par rapport aux autres bactéries présentes. Les compositions des différents échantillons sont détaillées dans le **Tableau 20**. Ici nous avons choisi de garder une concentration de *Francisella tularensis* subsp. *holarctica* LVS constante dans les mélanges de manière à évaluer l'effet de l'augmentation de la concentration des autres bactéries. Pour les mélanges aux rapports 1:199 et 1:499, les échantillons sont centrifugés et le culot bactérien est repris dans 1 ml de PBS avant l'immunocapture. Pour chaque préparation de mélange, un volume de 1 ml de préparation est conservé de manière à avoir un témoin du mélange avant l'immunocapture.

Une partie des spectres MALDI-TOF correspondants à cette expérience sont présentés sur la **Figure 40**. Ce sont des exemples de spectres obtenus pour les trois types d'échantillons, *Francisella tularensis* subsp. *holarctica* LVS pure, *E. coli* BL21 (DE3) pure et le mélange des deux bactéries avant et après l'immunocapture. Les scores moyens des différents échantillons sont présentés sur la **Figure 41**. Comme attendu, la bactérie est identifiée dans les spectres des échantillons de *Francisella tularensis* subsp. *holarctica* LVS purs à  $10^9$  CFU/ml avant et après l'immunocapture (109Témoin et 109Eluat). Il n'y a pas d'identification de *Francisella tularensis* subsp. *holarctica* LVS dans les échantillons contenant uniquement *E. coli* BL21 (DE3), avant et après immunocapture. Il est important de préciser que pour *E. coli* BL21 (DE3), il n'y a plus de profil protéique observé sur le spectre après l'immunocapture avec les billes couplées avec des anticorps spécifiques à *Francisella tularensis* subsp. *holarctica* LVS. Comme nous pouvons le voir sur la **Figure 41**, il n'y a pas d'identification de la bactérie dans les témoins de mélange. En revanche, une fois que l'immunocapture a eu lieu, les identifications sont possibles pour certains rapports de concentration. Pour le rapport 1:49, tous les éluats permettent une identification de la bactérie avec un score moyen de  $0,42 \pm 0,19$ . Pour le rapport 1:99, 5 éluats sur 9 permettent une identification, puis 4 sur 9 pour les rapports 1:199 et 1:499. Cette variabilité dans les résultats peut s'expliquer par le fait que la solution de *E. coli* BL21 (DE3) à  $10^{10}$  CFU/ml est particulièrement visqueuse. Au moment de l'incubation des billes avec les bactéries, la cinétique de fixation des *Francisella* doit être modifiée par cette viscosité. Une durée plus longue d'incubation pourrait être envisagée dans le cas de solutions à forte viscosité.

Tableau 20 : Méthode de préparation des mélanges bactériens.

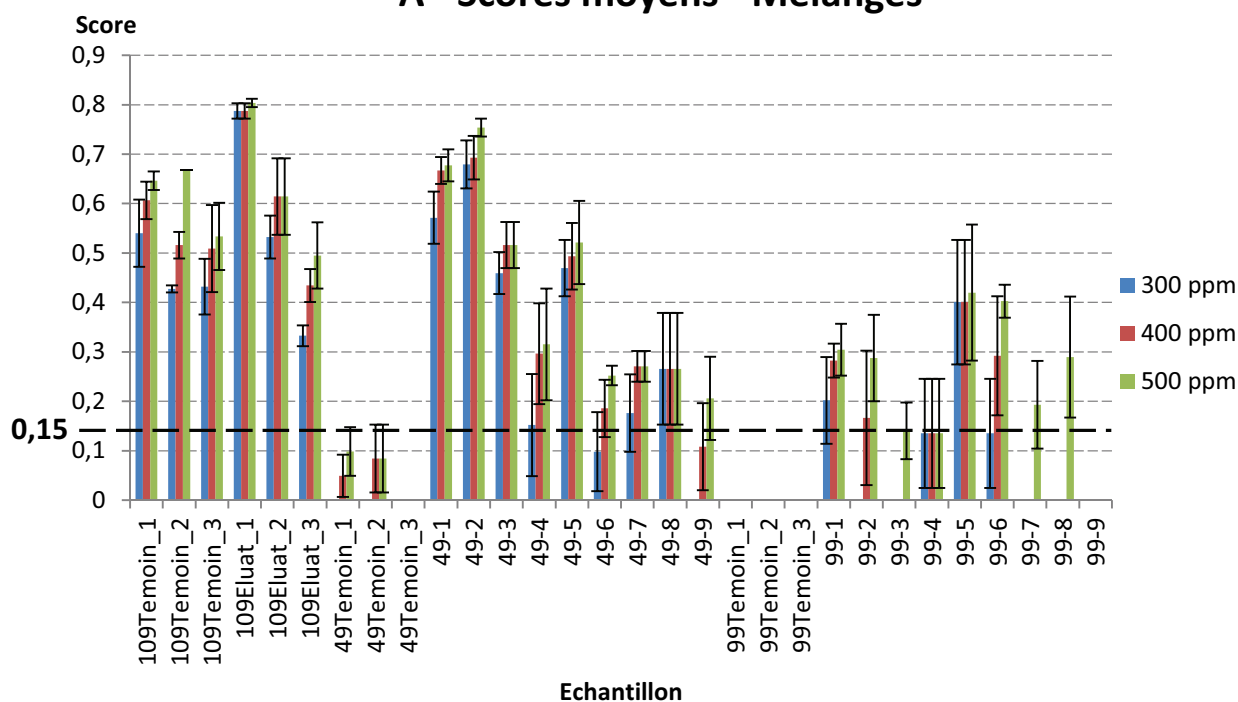
Jour 1	Jour 2	Jour 3	Rapport <i>Francisella</i> : <i>E.coli</i>	Volume et concentration initiale <i>Francisella</i>	Volume et concentration initiale <i>E.coli</i>	Concentration avant immunocapture <i>Francisella</i>	Concentration avant immunocapture <i>E.coli</i>
109Eluat_1	109Eluat_2	109Eluat_3	1:0	1 ml à 10 <sup>9</sup> CFU/ml	-	10 <sup>9</sup> CFU/ml	-
49.1	49.4	49.7	1:49	20 µl à 10 <sup>10</sup> CFU/ml	980 µl à 10 <sup>10</sup> CFU/ml	2.10 <sup>8</sup> CFU/ml	10 <sup>10</sup> CFU/ml
49.2	49.5	49.8					
49.3	49.6	49.9					
99.1	99.4	99.7	1:99	10 µl à 10 <sup>10</sup> CFU/ml	990 µl à 10 <sup>10</sup> CFU/ml	10 <sup>8</sup> CFU/ml	10 <sup>10</sup> CFU/ml
99.2	99.5	99.8					
99.3	99.6	99.9					
199.1	199.4	199.7	1:199	10 µl à 10 <sup>10</sup> CFU/ml	1990 µl à 10 <sup>10</sup> CFU/ml	10 <sup>8</sup> CFU/ml	2.10 <sup>10</sup> CFU/ml
199.2	199.5	199.8					
199.3	199.6	199.9					
499.1	499.4	499.7	1:499	10 µl à 10 <sup>10</sup> CFU/ml	5 ml à 10 <sup>10</sup> CFU/ml	10 <sup>8</sup> CFU/ml	5.10 <sup>10</sup> CFU/ml
499.2	499.5	499.8					
499.3	499.6	499.9					
EC109Eluat_1	EC109Eluat_2	EC109Eluat_3	0:1	-	1 ml à 10 <sup>9</sup> CFU/ml	-	10 <sup>9</sup> CFU/ml



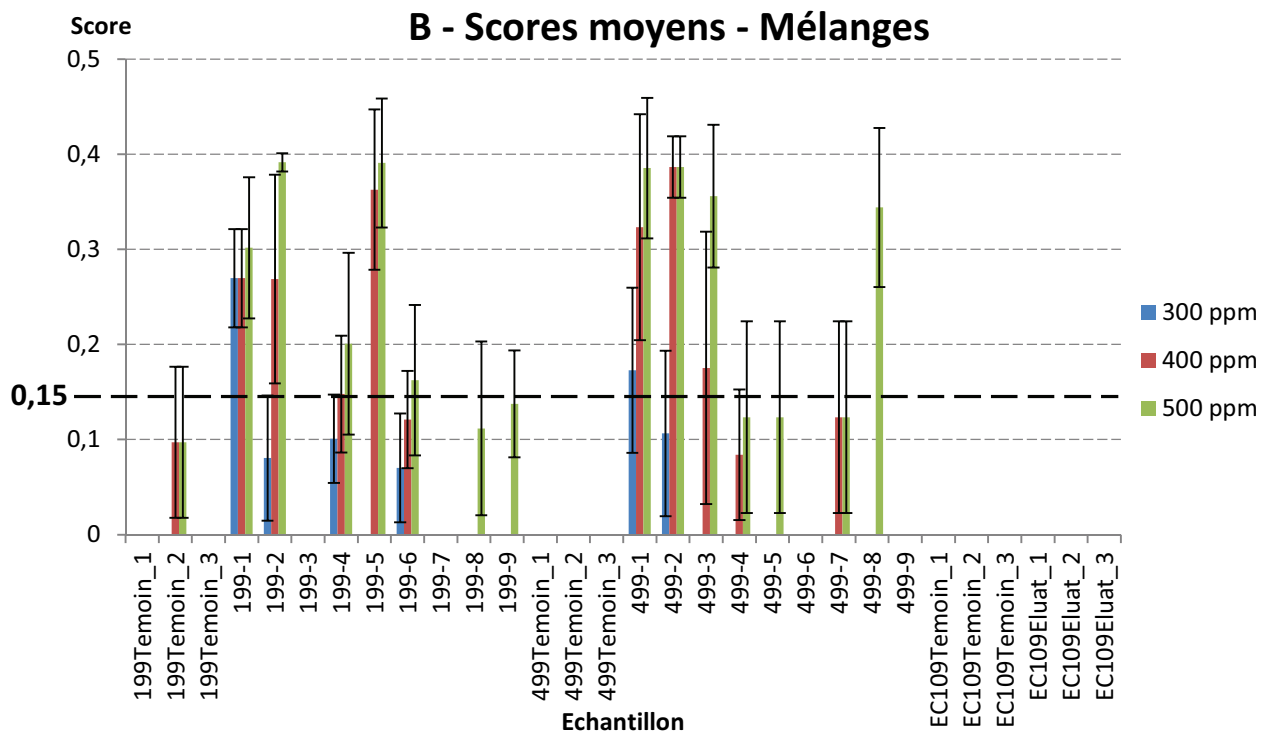
**Figure 40 :** Spectres MALDI-TOF des solutions de bactéries avant immunocapture (IC), colonne de gauche, et après immunocapture, colonne de droite. Présentés dans la gamme de  $m/z$  de 2 000 à 15 000 en fonction de leur intensité relative (IR).



### A - Scores moyens - Mélanges



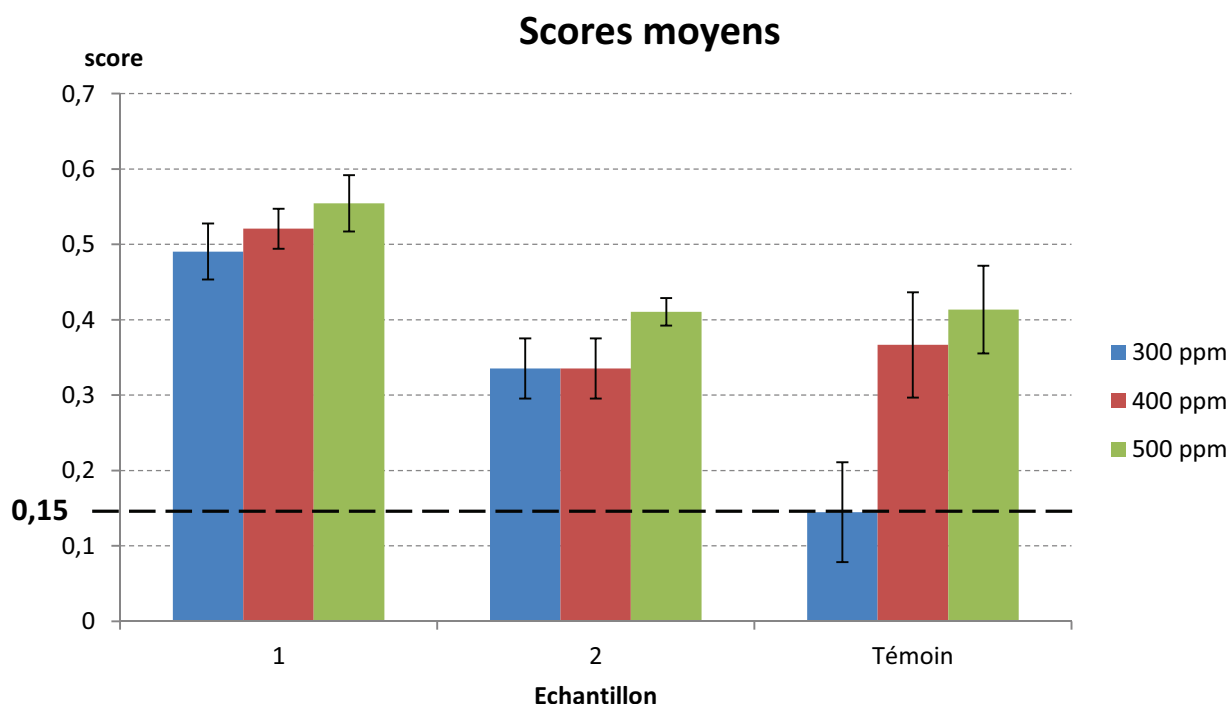
### B - Scores moyens - Mélanges



**Figure 41 : A :** Scores moyens du témoin positif (109Temoin et 109Eluat) et des mélanges de rapports 1:49 et 1:99. **B :** Scores moyens du témoin négatif (EC109Temoin et EC109Eluat) et des mélanges de rapport 1:199 et 1:499. Erreurs : en bleu, 300 ppm ; en rouge, 400 ppm ; en vert, 500 ppm.

### 8.3 Echantillons dilués

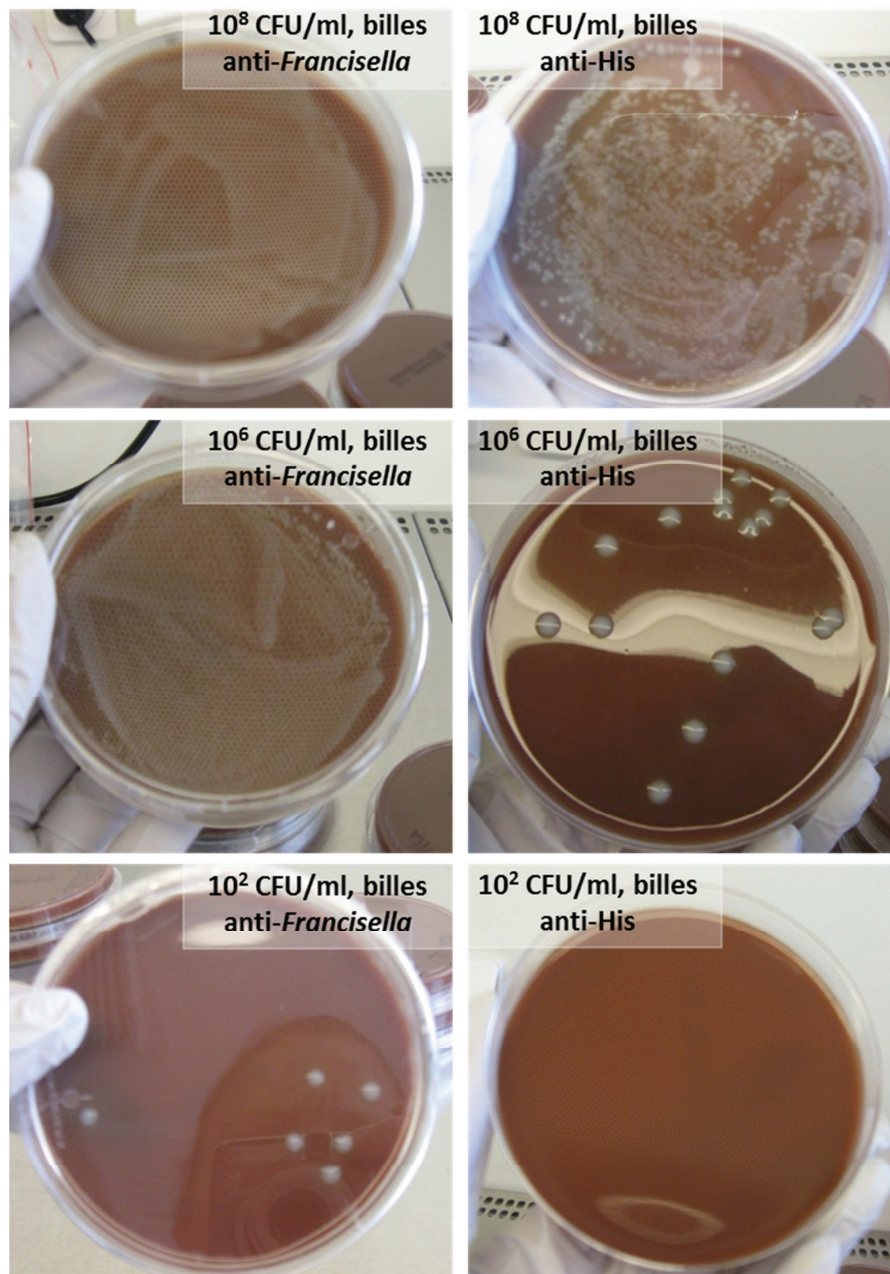
L'immunocapture en grand volume et à faible concentration de bactéries a aussi été expérimentée. Nous avons préparé des solutions de 500 ml de *Francisella tularensis* subsp. *holarctica* LVS aux concentrations de  $10^6$  et  $10^7$  CFU/ml dans du PBS. Le volume de 500 ml est ramené à 1,5 ml en utilisant une pompe péristaltique couplée à un filtre à  $0,2\ \mu\text{m}$  CellTrap™ CT22. La solution concentrée subit une immunocapture de *Francisella tularensis* subsp. *holarctica* LVS et ce sont les résultats de cette immunocapture qui sont présentés sur la **Figure 42**. Alors que sans immunocapture aucun spectre MALDI-TOF ne serait obtenu, la pré-concentration avec le filtre pour obtenir un volume de 1,5 ml et l'immunocapture nous ont permis d'identifier la bactérie aux différentes concentrations testées. Pour l'échantillon dont la concentration initiale était de  $10^6$  CFU/ml, la bactérie est identifiée par l'algorithme avec un score moyen de  $0,55 \pm 0,06$  à 500 ppm d'erreur. Pour l'échantillon dont la concentration initiale était  $10^7$  CFU/ml, la bactérie est identifiée par l'algorithme avec un score moyen de  $0,55 \pm 0,06$  à 500 ppm d'erreur. Ce système de pré-concentration avec un filtre CellTrap™ apparaît comme un très bon moyen de traitement des bactéries très diluées lorsqu'un grand volume d'échantillons est disponible. En revanche, ce dispositif ne peut être utilisé que sur des solutions à viscosité faible (eau par exemple) car des échantillons visqueux comme du lait, peuvent colmater les filtres.



**Figure 42** : Scores moyens de la solution à  $10^6$  CFU/ml (Echantillon 1) et à  $10^7$  CFU/ml après pré-concentration au Cell Trap et immunocapture. Le témoin est une solution de *Francisella tularensis* subsp. *holarctica* LVS à  $10^9$  CFU/ml. Erreurs : en bleu, 300 ppm ; en rouge, 400 ppm ; en vert, 500 ppm.

## 9 Remise en culture de bactéries après immunocapture

Comme nous avons pu le voir dans les paragraphes précédents, la sensibilité de l'appareil MALDI-TOF utilisé est faible pour l'étude de bactéries. Dans l'objectif de déterminer si *Francisella tularensis* subsp. *holarctica* LVS est réellement présente ou non dans un échantillon et ce, quelle que soit sa concentration, nous ne pouvons pas nous contenter de constater la présence ou l'absence de signal sur les spectres MALDI-TOF. En effet, dès que la quantité initiale de bactéries est inférieure à  $10^8$  CFU, même après l'immunocapture il est impossible de les détecter sur les spectres MALDI-TOF. Nous proposons donc une stratégie innovante pour éviter des faux négatifs (pas de spectres mais présence de bactéries), en faisant une remise en culture sur des boîtes de pétri des billes magnétiques ayant capturé des bactéries, suivi d'une analyse après culture des colonies par MALDI-TOF. L'élution des bactéries vivantes étant destructrice, cette étape n'est pas réalisée avant la mise en culture des billes. Ainsi, la remise en culture des bactéries immunocapturées permet d'amplifier le signal à mesurer. Lors d'essais préliminaires, nous nous sommes aperçus que mettre la totalité des billes (80  $\mu$ l) sur les boîtes de gélose empêchait la pousse de *Francisella tularensis* subsp. *holarctica* LVS. Nous avons donc essayé de mettre en culture 10  $\mu$ l et 1  $\mu$ l de billes avec d'excellents résultats. Dans les résultats qui vont être présentés ci-dessous, les immunocaptures ont été réalisées avec 1  $\mu$ l de billes magnétiques couplées avec des anticorps FT100R ou avec des anticorps anti-His, non spécifiques à *Francisella tularensis* subsp. *holarctica* LVS. A partir de bactéries en solutions à DO = 1 (à 600 nm) dans du PBS (soit environ  $10^9$  CFU/ml), nous avons préparé une gamme de dilution de dix en dix (de  $10^9$  à 10 CFU/ml). Dans ce cas, pour préserver au maximum l'intégrité des bactéries, les trois lavages ont été réalisés avec 100  $\mu$ l de PBS. Au final, le volume de 1  $\mu$ l de billes est resuspendu dans 100  $\mu$ l de PBS avant d'être déposé sur les boîtes de gélose. Des photos des boîtes de gélose après 4 jours de culture à 37°C sont présentées en **Figure 43**.

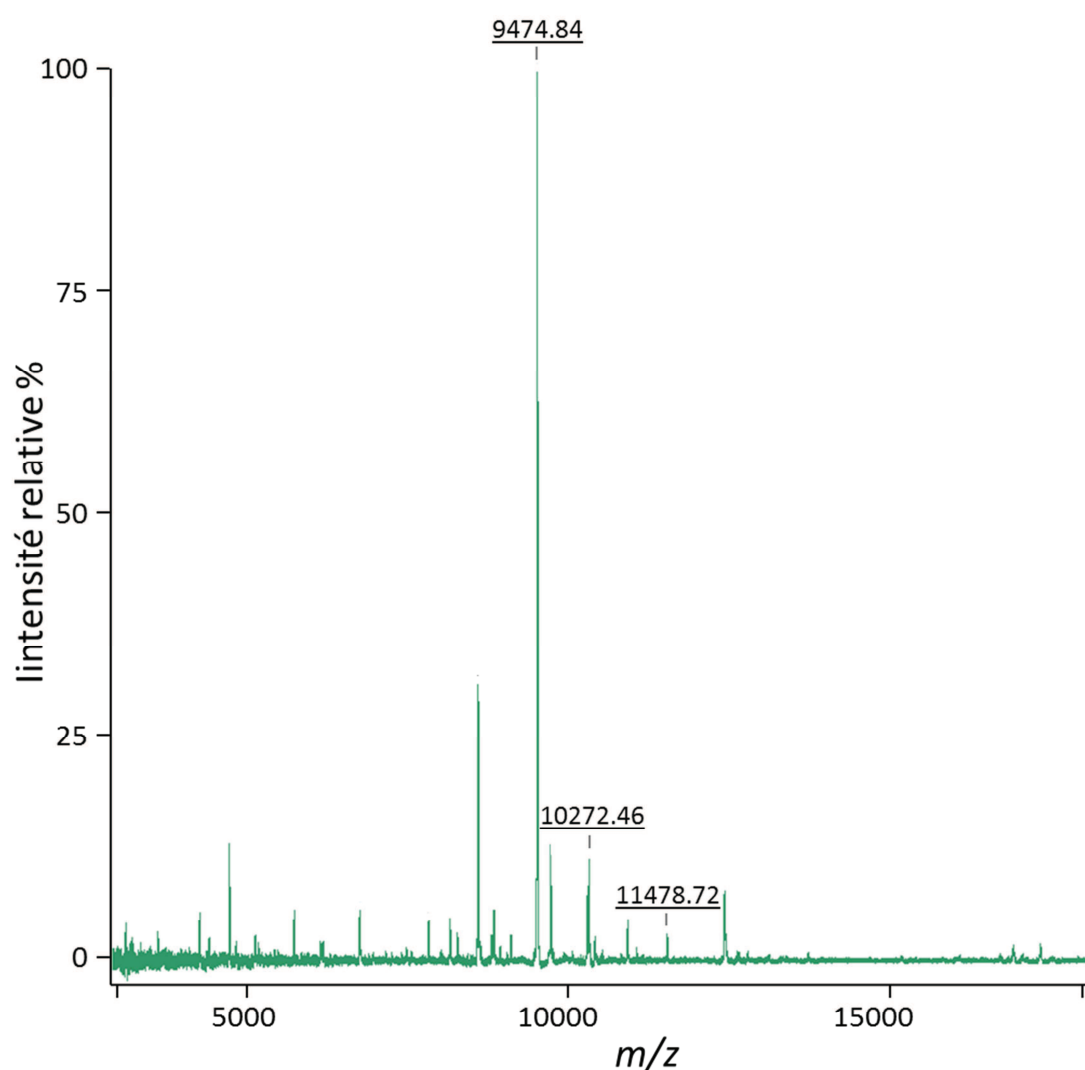


**Figure 43 :** Photos des boîtes de gélose ensemencées avec 1  $\mu$ l de billes magnétiques diluées dans 100  $\mu$ l de PBS après immunocapture et après 4 jours à 37°C. La colonne de gauche présente les remises en cultures des billes anti-*Francisella* avec les concentrations de *Francisella*, avant immunocapture, de 10<sup>8</sup> CFU/ml, 10<sup>6</sup> CFU/ml et 100 CFU/ml. La colonne de droite présente les remises en culture de billes anti-His aux mêmes concentrations de *Francisella*, avant immunocapture.

Avec les billes anti-*Francisella* une pousse est observée jusqu'à la concentration de 100 CFU/ml. A 10 CFU/ml une seule boîte sur les 3 boîtes ensemencées présente une colonie. L'immunocapture suivi de la remise en culture permet donc d'avoir une sensibilité de l'ordre de 100 CFU/ml. Pour les billes anti-His, non spécifiques à *Francisella tularensis* subsp. *holarctica* LVS, une pousse est observée de 10<sup>9</sup> à 10<sup>6</sup> CFU/ml. Pour les concentrations inférieures, les boîtes sont vierges. Ceci montre qu'un bruit de fond existe pour les très fortes concentrations de bactéries lorsque les anticorps ne sont pas spécifiques. Un lavage plus intensif pourrait permettre d'éviter cet écueil. L'intérêt de la mise en culture sur boîte

est qu'il est possible d'isoler les bactéries cibles des bactéries du bruit de fond sous la forme de colonies distinguables, ce qui serait impossible par une remise en culture en milieu liquide. En observant les photos (**Figure 43**), nous pouvons constater que les colonies de *Francisella tularensis* subsp. *holarctica* LVS sont alors beaucoup plus grosses que lors d'un étalement sur boîte classique. Ceci s'explique par le fait que les billes s'agrègent autour des bactéries et les colonies peuvent être donc formées par un amas de bactéries et non une bactérie isolée.

Après cette mise en culture, nous avons pu vérifier qu'il était possible d'identifier *Francisella tularensis* subsp. *holarctica* LVS en réalisant une inactivation au TFA suivi d'un spectre MALDI-TOF pour observer les trois biomarqueurs identifiés dans la 1<sup>ère</sup> partie de résultats, comme le montre la **Figure 44**.



**Figure 44** : Spectre MALDI-TOF acquis sur la gamme de  $m/z$  de 3 000 à 20 000 d'une remise en culture d'une immunocapture de *Francisella tularensis* subsp. *holarctica* LVS inactivée au TFA. Une calibration interne est réalisée avec le Protein Calibration Standard I de Bruker.

## 10 Conclusions

Nous avons mis au point un protocole permettant d'enrichir et de purifier un microorganisme cible dans diverses matrices à l'aide de billes magnétiques couplées à des anticorps.

### 10.1 Les inconvénients de cette méthode

#### 10.1.1 Limite de détection

Comme nous avons pu le voir, la limite de détection du MALDI-TOF est faible lorsqu'il s'agit de détecter des organismes entiers ( $10^6$  bactéries déposées au minimum), ce qui implique que les échantillons doivent être très concentrés à l'état initial ou qu'une étape de pré-concentration de grands volumes d'échantillons est nécessaire pour obtenir du signal. Une alternative serait de faire l'immunocapture directe de certaines protéines comme les biomarqueurs que nous avons identifiés dans la 1<sup>ère</sup> partie de résultats et éventuellement de faire une digestion enzymatique ce qui permettrait de gagner encore plusieurs ordres de grandeur en termes de sensibilité. Ceci en contrepartie nécessiterait plus de temps de préparation.

#### 10.1.2 Manque de discrimination niveau sous-espèce

Les anticorps choisis pour cette étude sont des anticorps visant le LPS de *Francisella tularensis* subsp. *holarctica* LVS. Ce LPS est commun avec le LPS des bactéries de la sous-espèce la plus virulente, *Francisella tularensis* subsp. *tularensis*. Cette méthode permet donc d'enrichir sans distinction les deux sous-espèces de bactéries qui n'ont pas la même virulence, il est donc important de pouvoir les différencier. En se basant sur le choix des biomarqueurs que nous avons définis dans la première partie de résultats, nous sommes en mesure de distinguer ces deux sous-espèces.

Cependant, nous avons pu voir dans ce chapitre que lorsque la concentration de bactérie est à la limite de détection, un signal est présent sur les spectres MALDI-TOF confirmant la présence d'une bactérie, en revanche les pics des biomarqueurs que nous avons définis dans la première partie de résultats ne sont pas toujours tous présents empêchant de distinguer de manière rapide quelle est la sous espèce présente. Deux solutions sont possibles pour pallier ce problème : i) faire une culture de la bactérie afin d'avoir une concentration suffisante pour observer les différents biomarqueurs, ii) posséder le spectre de référence des deux sous-espèces pour pouvoir comparer le spectre inconnu avec les deux spectres de références à l'aide d'un algorithme.

### 10.2 Les avantages de cette méthode

#### 10.2.1 Rapidité

Cette méthode présente l'avantage d'être rapide. En effet, l'immunocapture magnétique prend environ 90 minutes et l'acquisition du spectre MALDI-TOF se fait en quelques secondes. De plus, la culture peut être évitée si l'échantillon de départ est assez concentré et surtout, il permet d'isoler la bactérie cible dans le cas d'un mélange bactérien sans avoir à

passer par une procédure de culture et d'isolement qui peut être très coûteuse en temps. Si le microorganisme n'est pas cultivable ou difficile à cultiver, cela permet de s'affranchir de cette étape.

#### 10.2.2 Robustesse

Le couplage des anticorps sur les billes est extrêmement reproductible, avec un rendement de couplage de 90 %. De plus, le protocole d'immunocapture est robuste car les résultats sont reproductibles au cours du temps.

#### 10.2.3 Automatisation

Ce protocole est court lorsqu'il y a peu d'échantillons à traiter, en revanche lorsqu'il y a plus d'une dizaine d'échantillons cela devient très fastidieux. Nous avons donc envisagé la possibilité d'automatiser cette procédure avec un automate permettant de traiter simultanément jusqu'à 12 échantillons de 1 ml ou 6 échantillons de 5 ml. Le protocole automatisé sur le KingFisher<sup>TM</sup> Duo (Thermo Scientific) a été mis au point. Une phase de tests de reproductibilité devra être menée pour évaluer cet appareillage. L'automatisation permettrait également d'optimiser de nouveaux paramètres tels que le temps d'incubation ou le volume de billes optimum et de valider plus rapidement ce protocole pour d'autres microorganismes.

#### 10.2.4 Culture

La mise en culture des bactéries captées par les billes magnétiques permet à la fois d'éviter d'avoir des faux-négatifs et d'enrichir des bactéries qui seraient présentes en trop faible quantité pour une bonne détection. D'un point de vue plus général, il est intéressant de constater que les bactéries captées par les billes magnétiques restent vivantes car cela pourrait permettre, dans le cadre d'études environnementales, d'isoler de nouvelles souches de ce type de bactéries et de les mettre en culture pour pouvoir les étudier afin de s'en prémunir dans le cadre de recherches sur les pathogènes émergents.

# Discussion





# Discussion

## 1 Apports et complémentarité des deux méthodes développées.

Ce projet de thèse a permis la mise au point de deux méthodes complémentaires pour la détection et l'identification de *Francisella tularensis*. La première méthode, détaillée dans la première partie a mis en évidence que l'identification de trois protéines sur un spectre de masse MALDI-TOF d'une souche de *Francisella* permet de déterminer son espèce, éventuellement sa sous-espèce et donc sa virulence. La seconde partie du travail de thèse a permis de montrer qu'il est possible de concentrer et de purifier *Francisella tularensis* dans des matrices alimentaires telles que de l'eau minérale ou du lait, mais aussi à partir de mélanges de bactéries, même en proportion très désavantageuse. Ce protocole est basé essentiellement sur des billes magnétiques couplées à des anticorps dirigés contre le lipopolysaccharide de *Francisella tularensis* subsp. *holarctica* LVS. La détection par MALDI-TOF permet d'avoir une réponse immédiate quant à la présence ou non d'une bactérie, et l'algorithme mis au point permet de valider l'identification de celle-ci. La méthode de remise en culture des billes magnétiques présentée dans la seconde partie permet d'éviter un résultat faux-négatif lié à un manque de sensibilité de l'appareil de spectrométrie de masse utilisé. Cette étape implique néanmoins un délai d'attente supplémentaire pour une réponse définitive.

Ces différentes méthodes de détection et d'identification peuvent être utilisées de manière complémentaire afin d'obtenir un diagnostic rapide sur la présence et la virulence de souches pathogènes dans l'environnement. Ces méthodes sont génériques car elles peuvent être appliquées à d'autres microorganismes, et notamment à ceux du programme GEFREASE, *Bacillus anthracis*, *Yersinia pestis* et *Vibrio cholerae*.

## 2 Application à d'autres microorganismes de la menace

Le programme GEFREASE vise à développer des technologies pour l'identification de *Francisella tularensis* mais aussi de plusieurs autres microorganismes : *Bacillus anthracis*, *Yersinia pestis*, *Vibrio cholerae* et pox virus. La méthode d'immunocapture sera adaptée à chacun de ces microorganismes pour ensuite les détecter et les identifier par spectrométrie de masse MALDI-TOF, une fois que les anticorps spécifiques de chacun des microorganismes, en cours de production, seront disponibles. Pour les autres bactéries, *Bacillus anthracis*, *Yersinia pestis* et *Vibrio cholerae*, la mise au point réalisée sur *Francisella tularensis* devrait permettre un transfert rapide de technologie. A ce propos, l'équipe de Becher (CEA-Saclay) a récemment publié des études très intéressantes sur l'identification et la validation de marqueurs spécifiques de spores de *Bacillus anthracis* par une approche protéogénomique et sur la détection de *Yersinia pestis* par immunocapture de cellules entières (Chenau et al., 2014a; Chenau et al., 2014b). Notamment, cette équipe disposant d'anticorps dirigés contre la protéine BdA (abondante à la surface des spores), a développé un protocole d'immuno-

enrichissement des spores, suivi d'une double protéolyse des protéines extraites par 80 % d'acide trifluoroacétique, puis un suivi par MRM ciblant les isoformes de la protéine SASP-B (Chenau et al., 2011). Dans ce cas-là, la limite de détection a été estimée à 7 000 spores. L'inconvénient de leur méthodologie centrée sur l'analyse de peptides est toutefois la nécessité de réaliser des digestions des protéines par des protéases, ce qui engendre forcément des temps de réponse plus long, mais aussi l'utilisation d'un matériel sophistiqué, *i.e.* nanoHPLC couplée à un spectromètre de masse en tandem de triple quadripôle. Mais, le plus grand nombre de génomes séquencés actuellement disponibles joue également en défaveur de cette approche trop ciblée. Les isoformes de la protéine SASP-B sont identiques en termes de séquence pour les organismes *B. cereus biovar anthracis C1* et *B. thuringiensis BGSC 4CC1* par exemple. L'équipe a alors été proposée une analyse multi-paramétrique pour proposer de nouveaux biomarqueurs pertinents pour la détection spécifique de spores de *B. anthracis* (Chenau et al., 2014a). La limite de détection de cette nouvelle méthodologie également basée sur l'analyse de transitions spécifiques (dans ce cas, 28 correspondant à 5 marqueurs protéiques) n'est pas rapportée. La méthodologie est efficace sur des mélanges de spores. Dans le cas d'une analyse de bactéries, la récente étude proposée par Chenau et al. (Chenau et al., 2014b) est riche d'enseignements. La détection de *Yersinia pestis* se fait à l'aide d'une immunocapture par des anticorps dirigés contre la protéine Pla, suivi d'une extraction de protéines par 80 % d'acide trifluoroacétique, protéolyse à la trypsine, et suivi par SRM de transitions spécifiques à trois biomarqueurs, *i.e.* toxine murine, activateur plasminogène et pesticine. La limite de détection est de 20 000 CFU/ml pour cette méthode qui peut s'appliquer sur de l'eau, ou une matrice plus complexe telle du lait (Chenau et al., 2014). L'avantage lié à l'analyse ciblée de peptides est un gain de sensibilité indéniable par rapport aux approches top-down, mais le corollaire est un temps de traitement de l'échantillon plus élevé. Leurs résultats confirment en tout cas la validité de nos approches et la possibilité de les appliquer à différents microorganismes.

En revanche pour le pox virus une étape d'optimisation supplémentaire de notre méthodologie devrait être nécessaire. En effet, les virus possèdent peu de protéines et d'après les électrophorèses en gel 1D produits par l'équipe de Yoder pour les protéines du noyau et de la membrane, très peu de protéines présentent un poids moléculaire inférieur à 20 kDa, limite correspondant à la gamme de masse que nous étudions lors de l'analyse MALDI-TOF (Yoder et al., 2006). Il semble donc incontournable d'introduire une étape de digestion protéolytique pour pouvoir espérer une sensibilité suffisante pour l'identification de pox virus avec des spectres MALDI-TOF. Ceci ajoutera une étape supplémentaire au protocole que devra effectuer la personne en charge de la confirmation et donc également un temps d'attente supplémentaire pour la confirmation de l'identification.

Pour les toxines prévues dans le projet, *i.e.* ricine, toxines botuliniques de type A, B, C, D, E et F, entérotoxines A et B, abrine, les anticorps sont développés au RKI et la méthode d'enrichissement a déjà été expérimentée par cette équipe, également avec des billes magnétiques sur la ricine, l'entérotoxine B et les toxines botuliques A et B avec des limites

de détection de l'ordre de 500 fmol pour chaque toxine (Kull et al., 2010). Le protocole d'immunocapture et d'élution étant très semblable à notre protocole d'enrichissement, un protocole commun pour les bactéries, les virus et les toxines pourrait être envisagé.

L'aspect de quantification des cibles n'est pas pris en compte dans le projet GEFREASE mais comme le montrent les articles de Cheneau *et al.* sur les marqueurs de *Bacillus anthracis* et sur l'immunocapture de *Yersinia pestis* (Cheneau et al., 2014a; Cheneau et al., 2014b), il est possible de développer des méthodes de quantification par SRM multiplexé en ESI-MS/MS pour suivre les différents marqueurs d'un microorganisme. Cela présenterait un intérêt de pouvoir quantifier les bactéries ou les toxines notamment dans le cadre du suivi de décontamination par exemple. Le développement de peptides ou protéines isotopiquement différenciables, tels que les standards PSAQ proposés par l'équipe de Brun (CEA-Grenoble), permettent des dosages extrêmement précis, dits quantification absolue. A propos de *Yersinia pestis*, l'approche ciblée par LC-SRM permet d'obtenir une limite de détection de  $2.10^4$  CFU/ml dans des matrices alimentaires, ce qui représente une différence de 4 log avec la méthode décrite dans ce manuscrit et qui est du même ordre de grandeur que les tests ELISA. Nous pouvons envisager d'abaisser cette limite par des spectromètres de masse de nouvelle génération, notamment ceux qui permettent d'accumuler des ions parents, avant déclenchement de la fragmentation, et mesures de transitions spécifiques.

### 3 Détection de nouveaux biomarqueurs à partir de l'exoprotéome

La première partie des résultats montre qu'il est possible grâce à la protéogénomique et aux progrès concernant le séquençage des génomes de prédire les masses attendues pour des protéines présentes chez toutes les souches, sous-espèces, espèces d'un genre à partir de l'étude d'une seule bactérie de ce genre. En utilisant le même principe il serait également possible de travailler sur l'exoprotéome des bactéries, c'est à dire sur les protéines qui sont sécrétées par les bactéries et donc sur des traces de présence de bactéries et non sur les bactéries elles-mêmes (Armengaud et al., 2012; Desvaux et al., 2009). En effet, les bactéries sécrètent un grand nombre de protéines pour s'adapter à leur milieu. Notamment, nous avons montré que ces exoprotéines sont en général très persistantes dans ces milieux extracellulaires car résistantes à la protéolyse (Durighello et al., 2014). Dans le cas de la bactérie marine *Phaeobacter strain DSM 17395*, une protéine fortement abondante dans le milieu de culture a été identifiée comme étant une toxine de type RTX qui de façon inattendue ne possédait que très peu de résidus basiques, et donc était récalcitrante à l'analyse protéomique classique utilisant la trypsin. Il serait donc possible de rechercher parmi les protéines sécrétées lesquelles sont des protéines spécifiques de telles ou telles bactéries afin de les rechercher dans des échantillons environnementaux ou médicaux. Ceci permettrait de s'affranchir de la présence en grande quantité de la bactérie ou de la lyse de certaines bactéries lors du prélèvement et du transport de l'échantillon puisqu'à l'heure actuelle notre démarche repose sur des protéines intracellulaires, qui peuvent être dégradées par des protéases lors du vieillissement de l'échantillon.

## 4 Détection environnementale

Comme nous avons pu le voir dans la seconde partie des résultats, il est possible de fortement concentrer des bactéries en faible concentration initialement grâce à une pré-concentration d'un grand volume de liquide avant d'effectuer l'immunocapture. Cette méthode pourrait être utilisée, avec des anticorps spécifiques d'un genre, d'une espèce ou d'une sous-espèce pour faire du screening environnemental de bactéries. En effet ce type d'immunocapture permettrait de collecter rapidement différentes souches d'un même genre de bactéries, et l'utilisation de biomarqueurs documentés permettrait d'effectuer de manière simple par spectrométrie de masse MALDI-TOF un rapprochement phylogénétique pour les souches collectées (Christie-Oleza et al., 2012).

Pour compléter l'approche d'immunocapture de cellules entières, nous pourrions également développer les anticorps correspondant aux trois biomarqueurs sélectionnés dans notre première approche. Il pourrait s'agir d'anticorps spécifiques à un genre particulier, en permettant de capturer de manière large un type de protéine, comme une protéine ribosomale, protéine dont la structure est très conservée. Le motif à reconnaître serait donc choisi dans une des zones conservées de la protéine, au sein de ce genre. Avec plusieurs types d'anticorps, il serait possible grâce aux masses précises obtenues au MALDI-TOF d'identifier des espèces ou des sous-espèces à partir d'échantillons environnementaux par exemple. Ce type de concept se heurte cependant à la production d'anticorps dirigés de façon spécifique vers des antigènes aussi particuliers qui peut se révéler coûteuse en temps de développement.

## 5 Le diagnostic médical : vers plus de sensibilité ?

Dans le diagnostic médical les contraintes sont différentes : les échantillons sont en faible quantité (prélèvement sanguin, échantillon d'urine, biopsie, etc.) et les réponses doivent être obtenues le plus rapidement possible. Il est donc nécessaire d'avoir des méthodes extrêmement sensibles et dont le processus est le plus rapide possible pour pouvoir traiter le patient le plus précocement possible et le plus justement, en évitant des antibiotiques à large spectre par exemple (Lavigne et al., 2013). A l'heure actuelle l'identification directe par MALDI-TOF sur des échantillons biologiques n'existe pas, des étapes de culture (hémocultures ou géloses au sang) ou de pré-concentration, filtration et séparation magnétique (urine) sont souvent nécessaires (Koehling et al., 2012; Stevenson et al., 2010). De nouvelles méthodes combinant spectrométrie de masse et PCR comme le système PLEX-ID (Abbott<sup>TM</sup>) qui associe PCR à large spectre sur des gènes de protéines ribosomales et des gènes de ménages à l'ESI-MS permettent d'identifier de manière semi-quantitative des organismes présents dans les échantillons (Emonet et al., 2010; Fournier et al., 2013). Le rapport  $m/z$  des amplicons est mesuré par ESI-MS permettant ainsi de déterminer la composition en bases nucléiques des amplicons et l'identification est réalisée grâce aux rapports  $m/z$  de plusieurs régions qui sont comparés avec des bases de données génomiques

(Ecker et al., 2008). Toutefois cette approche reste confidentielle, et les instruments ne sont pas commercialisés à l'heure actuelle.

Une approche ciblée utilisant ce principe pourrait être envisagée. En effet, à partir de biomarqueurs tels que ceux déterminés dans la première partie de résultats, il est possible de créer les amorces correspondant aux gènes des protéines recherchées de manière à les amplifier. L'amplification par PCR offre un gain de sensibilité important et la fabrication des amorces est également plus « simple » que la fabrication d'anticorps. Le biais de cette méthode est bien sûr que de l'ADN doit être présent dans l'échantillon.

## 6 Nouveautés instrumentales

Désormais, sur le marché sont disponibles de nouvelles configurations de spectromètre de masse qui permettent des analyses à très haut débit de milliers de peptides ou protéines et à haute résolution, voire à très haute résolution. Le spectromètre de masse en tandem Q Exactive HF proposé par la société Thermo Scientific est l'exemple type de ces nouveautés très performantes. L'analyse ciblée de plusieurs centaines de protéines peut être réalisée en mode PRM (Parallel Reaction Monitoring), ce qui devrait permettre de tester plusieurs marqueurs spécifiques et de couvrir l'ensemble des microorganismes de la classe 3.

Ces nouveaux appareils permettent d'acquérir un nombre de spectres MS/MS toujours plus importants et donc également d'identifier plus de peptides. De ce fait des approches sans *a priori* basées sur la technique shotgun permettant l'identification (Dworzanski et al., 2004; Jabbour et al., 2010) et la classification phylogénétique (Dworzanski et al., 2006; Dworzanski et al., 2010) des organismes sont en plein développement. Ces méthodes utilisent les cadres ouverts de lecture de l'ADN (Open Reading Frame, ORF) d'organismes dont le génome est entièrement séquencé pour les traduire en protéines qui forment des bases de données protéiques. Les protéines ainsi obtenues servent à créer une matrice d'occurrence qui contient les protéines de tous les organismes. A partir des protéines identifiées par spectrométrie de masse, cette matrice permet de classer un organisme par rapport aux autres organismes de la matrice d'occurrence en fonction de leur distance phylogénétique. L'équipe de Dworzanski a ainsi pu faire la classification phylogénétique de 20 souches de *Bacillus* (Dworzanski et al., 2010). L'identification est également possible en ne conservant dans la matrice d'occurrence que les protéines spécifiques de chaque organisme. Le travers de cette méthode est que la matrice ne contient pas tous les microorganismes donc un échantillon peut être attribué à son plus proche voisin dans la matrice et le nombre de génomes séquencés augmentant exponentiellement, le nombre de protéines spécifiques se restreint ce qui rend de plus en plus difficile l'identification. A partir de ces deux méthodes, le laboratoire d'accueil pour cette thèse a mis au point une nouvelle approche pour identifier les organismes à partir de données shotgun en utilisant l'ensemble des séquences de protéines contenues dans les bases de données protéiques. Les résultats obtenus permettent d'identifier des organismes même en mélange (Allain, 2014). Cette

méthodologie complètement nouvelle, sans aucun *a priori*, complémente idéalement les approches dites avec *a priori* puisque la méthodologie développée ne nécessite plus de savoir ce que l'on cherche ou de multiplier les expériences pour obtenir un résultat. Ces techniques présentent tout de même le désavantage d'être longues, puisqu'il faut digérer les échantillons par des protéases puis séparer les peptides par chromatographie en phase liquide avant l'analyse par spectrométrie de masse et enfin traiter informatiquement les résultats, et coûteuses car des appareillages dernier cri sont pour l'heure nécessaires.

En conclusion, l'émergence de la méthodologie « whole-cell MALDI-TOF » introduite désormais à grande échelle dans les laboratoires d'analyses hospitaliers a conduit à une révolution du diagnostic microbiologique. Notre étude contribue au développement de cette méthodologie pour l'identification d'agents pathogènes de classe 3 dans le cadre de programmes de lutte contre le bioterrorisme, en proposant les bases d'une méthodologie robuste, pouvant être transférée à de nombreux agents.

# Matériels et méthodes





# Matériels et méthodes

La section Matériels et Méthodes est organisée en deux parties, correspondant aux deux thèmes traités dans la partie Résultats. Certains protocoles sont identiques, ils seront décrits une seconde fois pour une lecture simplifiée des protocoles. Sauf contre-indications, les réactifs sont fournis par Sigma-Aldrich.

## 1 Recherche de biomarqueurs

### 1.1 Culture des bactéries

*Francisella tularensis* subsp. *holarctica* LVS est mis en culture sur des boîtes de gélose chocolat et PolyViteX (PVX) (bioMérieux). Après 48 h de culture à 37°C, des précultures de 4 ml de milieu liquide BHI (Brain Heart Infusion, bioMérieux) et PolyViteX 1 % (bioMérieux) sont préparées en inoculant les bactéries cultivées sur plaques. Les précultures sont incubées 8 h à 37°C sous une agitation de 160 rpm dans un incubateur agitateur (INFORS HT minitron, Liebherr). A 39 ml de milieu liquide BHI + PVX 1% est ajouté 1 ml de la préculture pour 18 h de culture à 37°C sous une agitation de 160 rpm. Les bactéries sont récoltées par centrifugation à 2 500 g, pendant 10 min à 20°C une fois la phase stationnaire atteinte.

### 1.2 Méthodes d'inactivation.

L'inactivation est réalisée avec de l'acide trifluoroacétique (TFA). Deux protocoles sont possibles : à partir des cultures sur boîte de gélose ou à partir des cultures en phases liquides.

#### Culture sur boîte de gélose :

L'équivalent de deux boucles prélevées à l'anse stérile sur la boîte de gélose est mélangé à 100 µl de TFA 80 %. Le mélange est incubé 30 min à 4°C sous une agitation de 180 rpm sur un Thermomixer® (Eppendorf). Puis 150 µl d'acétonitrile (ACN) à 67% sont ajoutés avant de centrifuger 5 min à 13 000 g à température ambiante (TA). Le surnageant contenant les protéines solubles est conservé à -80°C avant validation de l'inactivation.

#### Culture en milieu liquide :

Les bactéries humides sont pesées après leur récolte, le protocole suivant est adapté en fonction de la masse de bactéries obtenue. Pour 10 mg de bactéries humides est ajouté 100 µl de TFA 80 %. Le mélange est incubé 30 min à 4°C sous une agitation de 180 rpm sur un Thermomixer®. Puis 150 µl d'ACN 67 % sont ajoutés avant de centrifuger 5 min à 13 000 g à température ambiante (TA). Le surnageant contenant les protéines solubles est conservé à -80°C avant validation de l'inactivation.

### 1.3 Validation de l'inactivation

Une inactivation est validée lorsqu'aucune colonie n'est observée sur boîte de gélose chocolat + PVX après 6 jours de cultures à 37°C. Un volume de 100 µl de surnageant obtenu par les méthodes d'inactivation précédemment citées est déposé sur une boîte de gélose chocolat + PVX et mis en culture à 37°C pour une durée de 6 jours.

### 1.4 Analyse en spectrométrie de masse

#### 1.4.1 MALDI-TOF

Les échantillons protéiques obtenus après l'inactivation au TFA sont dessalés et concentrés en utilisant des ZipTip<sub>C18</sub> (Millipore) suivant les recommandations du constructeur. Un volume initial de 20 µl est dessalé et concentré pour un volume final de 5 µl dans du TFA 0,1 %.

Les échantillons sont ensuite déposés sur une plaque en acier, appelée cible, (MTP 384 target plate polished steel, Bruker) avec un calibrant, Protein Calibration StandardI (Bruker) et la matrice, l'acide sinapinique (SA, Bruker). La matrice est préparée en saturant une solution de 50 % ACN/1,25 % TFA avec l'acide sinapinique. Le calibrant déposé avec l'échantillon permet d'augmenter la précision sur les masses en réalisant une calibration interne après l'acquisition des spectres au MALDI-TOF réalisés avec une calibration externe comme cité précédemment par Christie-Oleza *et al.* (Christie-Oleza et al., 2013; Christie-Oleza et al., 2012). Le standard est composé de quatre protéines (cytochrome C, insuline, myoglobine et ubiquitine I) qui sont vus monochargés aux  $m/z$  12 360,97, 5 734,51, 16 952,30 et 8 565,76 respectivement et deux d'entre elles, le cytochrome C et la myoglobine sont vues dichargés aux  $m/z$  6 180,99 et 8 476,65 respectivement. Les dépôts sont réalisés de la manière suivante, un volume de 1 µl d'échantillon est mélangé avec 1,5 µl de matrice SA et 0,5 µl de calibrant dilué au dixième dans du TFA 0,1 %. Un volume de 1,5 µl de ce mélange est déposé sur la cible MALDI et le séchage a lieu à température ambiante.

Les spectres MS sont acquis sur une gamme  $m/z$  variant de 3 000 à 20 000 par un MALDI-TOF de type Biflex IV (Bruker) contrôlé par le logiciel FlexControl 2.4 (Bruker). Les spectres sont analysés avec le logiciel FlexAnalysis 2.4 (Bruker).

#### 1.4.2 NanoLC-MS/MS

##### a. Migration sur gel d'agarose

Les échantillons de bactéries inactivées au TFA sont également analysés par nanoLC-MS/MS. Pour cela, un volume de 15 µl du mélange LDS 3X et protéome inactivé au TFA de *F. tularensis* subsp. *holarctica* LVS est déposé dans un puits d'un gel NuPAGE® 4-12 % Bis-Tris (Invitrogen). Sur un deuxième puits, 10 µl de marqueurs de masse moléculaire (Novex® Sharp Pre-Stained Protein Standard, Invitrogen) sont déposés. Une migration longue en tampon de migration MES (acide 2-(N-morpholino)ethanesulfonique, Invitrogen) dans des cuves XCell SureLock Mini-Cell (Invitrogen) est effectuée à voltage constant de 200 V délivré par le générateur PowerPac 300 (Biorad) pendant 25 min. Après rinçage à l'eau milliQ, les

gels sont colorés au bleu de Coomassie avec un réactif prêt-à-l'emploi (SimplyBlue SafeStain, Invitrogen). Trois bandes de gel (équivalent à un volume d'environ 50 µL) contenant les protéines de masses inférieures à 20 kDa sont découpées avant d'être protéolysées.

### *b. Protéolyse in-gel*

Après chaque étape du protocole suivant, le surnageant est éliminé. Un volume de 200 µL de méthanol :  $\text{NH}_4\text{HCO}_3$  50 mM (1:1 v/v) est ajouté, le mélange est maintenu agité pendant 1 min à température ambiante (TA) à 500 rpm dans un Thermomixer®. Cette étape est répétée une fois. Le gel est déshydraté par l'ajout de 200 µL d'ACN: $\text{NH}_4\text{HCO}_3$  50 mM (1:1 v/v). Le mélange est agité pendant 5 min dans un Thermomixer® à 600 rpm. Un volume de 200 µL d'acétonitrile pur est ajouté aux morceaux de gel. Le mélange est agité pendant 1 min à 600 rpm. Le gel est séché au SpeedVac SPD121P (Savant) pendant 5 min. A cette étape le bleu de Coomassie et le SDS ont été éliminés. Les ponts disulfures sont ensuite réduits par un traitement au DTT 25 mM (dans du  $\text{NH}_4\text{HCO}_3$  50 mM) 20 min à 56 °C à 500 rpm. Le surnageant est éliminé. Les cystéines sont alkylées par de l'iodoacétamide 55 mM dans du  $\text{NH}_4\text{HCO}_3$  50 mM (20 min à température ambiante à l'abri de la lumière). L'iodoacétamide est alors éliminé et les morceaux de gels sont lavés à deux reprises par ajout de 200 µl d'eau milliQ. Le gel est déshydraté par ajout de 200 µl d'ACN: $\text{NH}_4\text{HCO}_3$  50 mM (1:1 v/v) et une incubation de 5 min à 600 rpm. Un volume de 200 µl d'acétonitrile pur est incubé pendant 1 min à 600 rpm et le gel est séché au SpeedVac SPD121P (Savant) pendant 5 min. La trypsine (Roche) est reprise dans de l'acide trifluoroacétique à 0,01% (v/v) pour une concentration finale de 0,1 mg/ml. Les protéines présentes dans le gel sont ensuite incubées dans 20 µl de solution enzymatique (10 ng/µl de trypsine (Roche) dans 0,01 % ProteaseMAX™ Surfactant (Promega) dans du  $\text{NH}_4\text{HCO}_3$  50 mM) sur de la glace pendant 20 min. L'excès de surnageant est éliminé et 50 µL de solution de ProteaseMAX™ Surfactant à 0,01 % dans du  $\text{NH}_4\text{HCO}_3$  50 mM sont ajoutés. La digestion dure 4 h à 37 °C. L'extrait peptidique (surnageant) est transféré dans un nouveau tube puis la trypsine est inactivée par ajout de 5 µl de TFA à 5 %.

### *c. Analyse en nanoLC-MS/MS*

Les analyses LC-MS/MS ont été réalisées sur un spectromètre de masse hybride LTQ-Orbitrap XL (ThermoFisher) couplé à un système LC UltiMate 3 000 (Dionex-LC Packings). Les échantillons ont été chargés et dessalés une nouvelle fois en ligne sur une précolonne phase inverse Acclaim Pepmap 100 C18 (5 µm taille de bille, 100 Å taille de pore, 5 mm x 300 µm) de LC Packings. Ils ont ensuite séparés sur une nanocolonne Acclaim Pepmap 100 C18 (3 µm taille de bille, 100 Å taille de pore, 15 cm x 75 µm) de LC Packings à un débit de 0.3 µL/min. Les peptides ont été séparés en utilisant un gradient de 60 min avec un solvant A (0.1%  $\text{HCOOH}$ ) et un solvant B (0.1 %  $\text{HCOOH}$ /80 %  $\text{CH}_3\text{CN}$ ) développé comme suit : 5-60 % B en 30 min, 60-90 % B en 1 min, 90 % B pendant 8 min, 90-5 % B en 1 min, et 5 % B pendant 20 min. Le spectre de masse en « full scan » est mesuré de  $m/z$  300 à 1 800. Le spectromètre de masse LTQ-Orbitrap XL a été utilisé en mode « data-dependent » avec la stratégie « TOP3 ». En résumé, un cycle de scan a été initié avec un scan complet de haute précision sur la

masse dans l'analyseur Orbitrap, suivi par des scans MS/MS dans la trappe ionique linéaire des trois ions précurseurs les plus abondants avec une exclusion dynamique des ions précédemment sélectionnés. Cette exclusion dynamique consiste en deux acquisitions de spectres MS/MS de l'ion le plus abondant pendant une période de 30 sec puis par l'exclusion de cet ion pour les fragmentations suivantes pendant les 60 sec suivantes. Le type d'activation utilisé est le CID avec une énergie de collision normalisée fixée à 30. La calibration utilisée dans l'analyseur Orbitrap pour la mesure des spectres FTMS est appelée « Lock mass » : elle consiste en une calibration interne à partir d'une liste de composés définis présents en permanence (ici, des polysiloxanes) et l'erreur sur la masse est inférieure à 2 ppm. Pour l'analyseur LTQ, il s'agit d'une calibration externe qui est effectuée mensuellement.

#### 1.4.3 Traitement des données MS

Les spectres MS/MS sont attribués à des peptides grâce au serveur MASCOT 2.3.02 (Matrix Science). C'est un puissant moteur de recherche (Perkins et al., 1999) qui permet l'identification de protéines dont les séquences sont contenues dans des bases de données, à partir des données de spectrométrie de masse. La base utilisée ici est la liste complète des séquences codantes (CDSs) de *Francisella tularensis* subsp. *holarctica* LVS (téléchargée depuis NCBI le 10/01/2012). Les recherches sont effectuées avec les paramètres suivants : peptides tryptiques avec un maximum de deux défauts de coupure peptidiques, une tolérance sur la masse de l'ion parent de 5 ppm et de 0, 5 Da sur les spectres MS/MS, la carbamidométhylation des cystéines comme modification fixe (+ 57,02 Da), et l'oxydation des méthionines comme modification variable (+ 15,99 Da). Les résultats MASCOT sont ensuite analysés par le logiciel IRMa 1.28.0 (Interprétation des Résultats MASCOT) (Dupierris et al., 2009), développé par l'équipe « EDyP » du CEA-Grenoble, qui permet de filtrer les résultats Mascot selon des critères établis par l'utilisateur. Le filtre élimine tous les peptides dont la probabilité qu'il s'agisse d'une fausse attribution (p-value) est supérieure à 0,05.

La probabilité d'erreur sur l'attribution de peptides est évaluée à moins de 0,1 % en utilisant la base de données en mode « decoy » (base artificielle créée en mélangeant de manière aléatoire les éléments de la base initiale). Une protéine est identifiée lorsqu'au moins deux peptides différents lui sont attribués.

#### 1.4.4 Analyse des séquences protéiques et nucléotidiques

L'algorithme BLAST (Basic Local Alignement Search Tool), (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) donne la possibilité de tester l'identité ou l'homologie de séquences d'acides aminés, de protéines entières ou de nucléotides avec des séquences connues de protéines ou de séquence d'ADN. Pour les protéines les recherches sont effectuées contre la base de données NCBI nr. Pour les séquences nucléotidiques les recherches sont réalisées sur chaque génome d'intérêt. Les génomes sont obtenus sur le site du NCBI où sont stockés les génomes microbiens complets ou en cours de séquençage ([http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial\\_taxtree.html](http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html)).

Les calculs des masses isotopiques et moyennes sont réalisés avec l'outil « Compute PI/Mw » disponible sur le portail de ressource bioinformatique ExPASy ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)).

## 2 Immunocapture magnétique

### 2.1 Anticorps

Tous les anticorps ont été développés et produits par le Laboratoire d'Ingénierie Cellulaire et Biotechnologie associé à ce travail de thèse (CEA/DSV/ibeb/LICB). Les anticorps FT100R, FT101R et FTC129Z sont anti-*Francisella tularensis*. Les anticorps anti-TMPP (TMP05) et anti-6-His sont utilisés comme témoins négatifs.

#### 2.1.1 Développement des anticorps

Une solution de *Francisella tularensis* subsp. *holarctica* LVS inactivées à la chaleur a été utilisée pour immuniser par injection intrapéritonéale 4 souris Biozzi et 3 rats en présence d'adjuvant de Freund complet. Avant l'immunisation il a été vérifié que les bactéries étaient correctement inactivées en ensemencant celles-ci sur des boîtes de gélose chocolat + PVX. L'inactivation est validée lorsqu'au bout d'une semaine de culture à 37°C aucune colonie n'était apparue sur la boîte. Chaque mois, les animaux ont reçu une nouvelle injection de l'immunogène et deux semaines après chaque injection, le sérum de chaque animal immunisé est prélevé et titré. Après la dernière injection (6 pour les souris, 3 pour les rats), les animaux dont la titration est la plus élevée sont sacrifiés et les lymphocytes B sont prélevés à partir de la rate. Les lymphocytes B sont ensuite fusionnés avec des myélomes murins P3-X63Ag8.653 (ATCC) avec les proportions 1 : 2 et 1 : 4 (myélomes : lymphocytes) pour les souris et les rats respectivement, en présence de polyéthylène glycol. Les hybridomes sont mis en milieu sélectif RPMI 1640 L-Glutamine avec supplément de HT (Hypoxantine, Thymidine) et d'aminoptérine (Gibco) et mis en culture dans des plaques 96 puits contenant 5 000 macrophages péritonéaux. Les plaques sont mises à incuber 10 jours à 37°C sous atmosphère humide à 5 % de CO<sub>2</sub>. Dix jours après la fusion, les surnageants de culture des hybridomes sont testés pour la présence d'anticorps spécifiques avec un test ELISA. Les tests ELISA sont réalisés sur des plaques où sont immobilisées des *Francisella tularensis* subsp. *holarctica* LVS inactivées. Les hybridomes positifs sont de nouveau testés sur des *Francisella tularensis* subsp. *holarctica* LVS vivantes et sur des *E. coli*. Les hybridomes positifs sont clonés en plaques 96 puits en utilisant la technique de la dilution limite, puis de nouveau testés par test ELISA sur bactéries vivantes.

#### 2.1.2 Test ELISA

Pour les tests ELISA des anticorps FT 101R sont immobilisés sur une microplaque 96 puits. Pour cela un volume de 120 µl d'anticorps à 20 µg/ml en tampon carbonate à pH=9 est distribué dans chaque puits avant une incubation de 4 h sous agitation. La microplaque est lavée par du PBS 0,1 % Tween-20. Les puits sont saturés avec 200 µl de PBS 0,5 % BSA. La microplaque est mise à incuber une nuit à 4°C. La microplaque est de nouveau lavée par du

PBS 0,1 % Tween-20 puis 50 µl d'anticorps FT100R biotinylé à 200 ng/ml en PBS 0,1 % BSA sont distribués dans chaque puits. Ensuite sont ajoutés 50 µl de la solution de bactéries à tester préparées à la concentration désirée dans du PBS 0,1% BSA (en général les dilutions sont au 1/50<sup>ème</sup>, 1/500<sup>ème</sup> et 1/5 000<sup>ème</sup>). La microplaque est laissée à incuber une nuit à 4°C. Ensuite, la microplaque est lavée à deux reprises par du PBS 0,1 % Tween-20 et 100 µl de streptavidine-péroxydase diluée au 1/50 000<sup>ème</sup> dans du PBS 0,1 % BSA est distribué dans chaque puit. Après 30 min d'agitation à 900 trs/min à TA, la microplaque est lavée 3 fois avec du PBS 0,1 % Tween-20. Un volume de 100 µl de TMB (3,3',5,5'-tétraméthylbenzidine) est distribué dans chaque puit et la microplaque est laissée à incuber 30 min à TA à l'abri de la lumière. Pour arrêter la réaction, 25 µl d'acide sulfurique 1 M (H<sub>2</sub>SO<sub>4</sub>) sont distribués dans chaque puit. La lecture de DO à 450 nm s'effectue par un spectrophotomètre à microplaques (Multiskan® Spectrum, Thermo Scientific).

## 2.2 Préparation des billes magnétiques

Plusieurs types de billes ont été testées avec différentes concentrations pour les couplages billes anticorps. Les résultats expliquant les choix des billes Dynabeads® MyOne™ Tosylactivated (Invitrogen) couplées avec les anticorps FT100R à 25 µg d'anticorps/mg de billes sont détaillés dans la partie résultat. Seul ce protocole de couplage est expliqué dans la partie Matériels et Méthodes.

Les billes sont conditionnées à la concentration suivante : 100 mg de billes/ml. Pour 4 mg de billes, prélever 40 µl de solution contenant les billes. Un portoir magnétique (DynaMag™-2, Invitrogen) adapté permet d'isoler contre une paroi du tube les billes magnétiques afin d'éliminer les différentes solutions sans perdre de billes. Les billes sont lavées à deux reprises avec 400 µl de tampon borate de sodium 0,1 M à pH=9,5 (6,183 g de H<sub>3</sub>BO<sub>3</sub> dans 1 l d'eau, ajuster le pH). Après élimination du dernier surnageant les billes sont resuspendues dans 33 µl de tampon borate de sodium 0,1 M à pH=9,5. Un volume de 25 µl de l'anticorps (FT100R) à une concentration de 4 mg/ml dans le tampon borate de sodium est ajouté, suivi d'un volume de 17 µl de sulfate d'ammonium 3 M (39.6 g de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> dans 100 ml de tampon borate de sodium, ajuster le pH), afin d'atteindre une concentration finale de 0,5 M. Les billes sont ensuite laissées à incuber une nuit à 37°C dans un Thermomixer® sous une agitation de 850 rpm.

Le surnageant est mis de côté de manière à mesurer la densité optique (DO) à 280 nm afin de déterminer la quantité d'anticorps couplés. La réaction de couplage est bloquée par l'ajout de 400 µl d'éthanolamine à 50 mM dans du PBS Tween 0,05 % (2 ml de PBS Tween 0,05% et 105 µl d'éthanolamine 1 M. Préparation du PBS : 0,26 g de NaH<sub>2</sub>PO<sub>4</sub> x H<sub>2</sub>O, 1,44 g de Na<sub>2</sub>HPO<sub>4</sub> x 2H<sub>2</sub>O et 8,78 g de NaCl dans 1 l d'eau, ajuster à pH=7,4). Les billes sont de nouveau laissées à incuber une nuit à 37 °C dans un Thermomixer® sous une agitation de 850 rpm.



Le surnageant est éliminé puis les billes sont lavées à deux reprises par 400 µl de PBS Tween 0,05 %. Après élimination du dernier surnageant les billes sont reprises dans 80 µl de PBS Tween 0,05 %. La solution de billes est stockée à 4°C avant d'être utilisée.

Les mesures de DO sont réalisées à l'aide d'un spectrophotomètre Cary 50 Probe (Varian), Les cuves, d'une contenance de 45 µl, sont en quartz avec un trajet optique de 3 mm (Hellma).

### 2.3 Les bactéries

Les souches de bactéries utilisées sont les suivantes : *Francisella tularensis* subsp. *holarctica* LVS et *Francisella philomiragia* ATCC 25015 détenues par le LICB et manipulables en laboratoire de niveau 2. *Francisella tularensis* subsp. *tularensis* CCUG 2112 est également détenue par le LICB et provient de l'équipe d'Edward R. Moore (Université de Göteborg, Suède). Cette bactérie doit être manipulée dans un laboratoire de niveau 3. *Escherichia coli* BL21(DE3) est détenue par le LBSP et peut être travaillée en laboratoire de niveau 1.

### 2.4 Culture des bactéries

*Francisella tularensis* subsp. *holarctica* LVS est mis en culture sur des boîtes de gélose chocolat et PolyViteX (PVX) (bioMérieux). Après 48 h de culture à 37°C, des précultures de 4 ml de milieu liquide BHI (Brain Heart Infusion, bioMérieux) et PolyViteX 1% (bioMérieux) sont préparées en inoculant les bactéries cultivées sur plaques. Les précultures sont incubées 8 h à 37°C sous une agitation de 160 rpm dans un incubateur agitateur (INFORS HT minitron, Liebherr). A 39 ml de milieu liquide BHI + PVX 1 % est ajouté 1 ml de la préculture pour 18 h de culture à 37°C sous une agitation de 160 rpm.

*Escherichia coli* BL21(DE3) est mis en culture sur des boîtes de milieu LB agar (bouillon lysogène) pendant 24 h à 37°C en incubateur. Une boucle de bactérie est prélevée sur la boîte de pétri et diluée dans 100 ml de milieu LB liquide dans un erlenmeyer à ailettes de 250 ml. La culture est mise en incubation 18 h à 37°C sous une agitation de 160 rpm dans un incubateur agitateur (INFORS HT minitron).

### 2.5 Préparation des bactéries

Les bactéries sont récoltées en répartissant le volume de culture dans deux Falcons de 50 ml en centrifugeant à 2 500 g, pendant 20 min à 20°C une fois la phase stationnaire atteinte. Après la récolte des bactéries, elles sont lavées à plusieurs reprises avec du PBS stérile afin d'éliminer le maximum de traces du milieu de culture. Un volume de 25 ml de PBS stérile pH=7,4 (cf. § 2.2 Préparation des billes magnétiques) est ajouté dans chaque Falcon pour resuspendre les bactéries. Les bactéries sont centrifugées à 2 500 g, pendant 15 min à 20°C et le surnageant est éliminé. Cette étape de lavage est répétée une deuxième fois.

A la fin des lavages les bactéries sont reprises dans le volume de PBS stérile nécessaire pour se placer à  $DO_{600nm}$  de 1 soit environ  $10^9$  CFU/ml pour *F. tularensis* subsp. *holarctica* LVS.



## 2.6 Comptage des bactéries

Pour avoir une estimation de la concentration des bactéries à chaque expérience, il est nécessaire de faire des comptages. Pour cela une dilution en cascade de 10 en 10 jusqu'à  $10^{-8}$  à partir des bactéries à  $DO_{600nm}$  de 1 dans du PBS est réalisée. Pour les dilutions à  $10^{-6}$ ,  $10^{-7}$  et  $10^{-8}$ , 3 fois 100  $\mu$ l de chaque dilution sont déposés sur des boîtes de géloses chocolat PVX. Après 4 jours d'incubation à 37°C, le nombre de colonies apparues est compté.

## 2.7 Immunocapture magnétique des bactéries

Les immunocaptures s'effectuent dans un volume d'échantillon de 1 ml. Les billes magnétiques, 4 mg de Dynabeads® MyOne™ Tosylactivated (Invitrogen) à 25  $\mu$ g d'Ac/mg de billes (80  $\mu$ l), sont lavées 3 fois avec 400  $\mu$ l de PBS stérile (cf. 2.2 Préparation des billes magnétiques). Un portoir magnétique (DynaMag™-2, Invitrogen) adapté permet d'isoler contre une paroi du tube les billes magnétiques afin d'éliminer les différentes solutions sans perdre de billes. Après l'élimination du dernier surnageant, un volume de 1 ml de l'échantillon à concentrer est ajouté. Le mélange est agité quelques secondes puis laissé à incuber 1 h à température ambiante sous agitation à 1 000 rpm sur un Thermomixer®. Ensuite le surnageant est éliminé et deux lavages avec 400  $\mu$ l de PBS stérile suivi d'un lavage par 400  $\mu$ l d'eau stérile sont effectués. Après l'élimination du dernier surnageant, les bactéries sont éluées avec 100  $\mu$ l de TFA 0,2 % en vortexant 10 secondes. A l'aide du portoir magnétique l'éluat est séparé des billes magnétiques et placé dans un nouveau tube. Les éluats sont en condition acide. Ils sont neutralisés en ajoutant 1 % du volume de l'éluat d'hydroxyde d'ammonium 25 %.

## 2.8 Inactivation des bactéries à la chaleur

Les solutions de bactéries sont chauffées à 90°C sous une agitation de 600 rpm pendant 1 h sur un Thermomixer®. Pour chaque série de tubes de bactéries inactivés, deux tubes de 1 ml de *F. tularensis* subsp. *holarctica* LVS à  $DO_{600nm}$  de 1 sont ajoutés. Ceux-ci serviront à valider l'inactivation des bactéries. A la fin du chauffage, les deux tubes sont centrifugés 5 min à 13 000 g. le surnageant est éliminé et le culot repris dans 100  $\mu$ l de PBS. Le volume contenu par les deux tubes respectivement est déposé sur deux boîtes de gélose chocolat PVX. Les boîtes sont mises à incuber à 37°C pendant 6 jours. L'inactivation est validée lorsqu'aucune colonie n'est apparue au bout des 6 jours de culture.

## 2.9 Remise en culture des bactéries

Après l'immunocapture et avant l'élution, il est possible de remettre les bactéries en culture. Pour cela, les bactéries sont mises en solution dans 100  $\mu$ l de PBS stérile et le volume est déposé sur une boîte de gélose chocolat PVX. Les boîtes sont mises en incubation à 37°C entre 4 et 6 jours.

## 2.10 Analyse en spectrométrie de masse

### 2.10.1 MALDI-TOF

Les échantillons protéiques obtenus après l'immunocapture et l'inactivation à la chaleur sont analysés par spectrométrie de masse MALDI-TOF.

Les échantillons sont déposés sur des plaques en acier, des cibles (MTP 384 target plate polished steel, Bruker) avec un calibrant, IVD Bacterial Test Standard (BTS, Bruker) et la matrice, l'acide  $\alpha$ -cyano-4-hydroxycinnamique (HCCA, Bruker). La matrice est préparée en saturant une solution de 50 % ACN/2,5 % TFA. Le calibrant est le profil protéique et peptidique typique d'*Escherichia coli* DH5 alpha auquel est ajouté plusieurs protéines afin de couvrir la gamme de  $m/z$  de 2 000 à 20 000. Les 7 protéines qui sont vues monochargées sont RS32, RS34, RS33meth, RL29, RS19 RNaseA et la myoglobine (vues aux  $m/z$  5 096,8, 5 381,4, 6 255,4, 7 274,5, 10 300,1, 13 683,2 et 16 952,3 respectivement). RL29 est également vu sous sa forme dichargée au  $m/z$  3 637,8. Les spectres MS sont acquis sur une gamme de  $m/z$  de 2 000 à 20 000 par un MALDI-TOF de type Biflex IV (Bruker) contrôlé par le logiciel FlexControl 2.4 (Bruker). Les spectres sont analysés avec le logiciel FlexAnalysis 2.4 (Bruker).

### 2.10.2 Validation des résultats

Les spectres sont traités par un algorithme (**Annexe 4**) créé à partir d'Octave-3.6.4. Basé sur une publication de Jarman *et al.* (Jarman et al., 2000), cet algorithme permet de comparer les spectres obtenus par spectrométrie de masse MALDI-TOF avec des spectres de références et de donner une probabilité de présence de la référence dans notre échantillon d'intérêt. Les scores obtenus et leur signification sont décrits dans le **Tableau 13**.



# Références bibliographiques



# Références bibliographiques

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198-207.
- Afonso, C., and Fenselau, C. (2003). Use of bioactive glass slides for matrix-assisted laser desorption/ionization analysis: Application to microorganisms. *Analytical Chemistry* 75, 694-697.
- Alanio, A., Beretti, J.L., Dauphin, B., Mellado, E., Quesne, G., Lacroix, C., Amara, A., Berche, P., Nassif, X., and Bougnoux, M.E. (2011). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry for fast and accurate identification of clinically relevant *Aspergillus* species. *Clinical Microbiology and Infection* 17, 750-755.
- Allain, F. (2014). Nouvelle méthode d'interprétation de données de spectrométrie de masse en tandem pour l'identification de microorganismes dans un échantillon complexe.
- Altelaar, A.F.M., Munoz, J., and Heck, A.J.R. (2013). Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* 14, 35-48.
- Anhalt, J.P., and Fenselau, C. (1975). Identification of bacteria using mass-spectrometry. *Analytical Chemistry* 47, 219-225.
- Armengaud, J., Christie-Oleza, J.A., Clair, G., Malard, V., and Duport, C. (2012). Exoproteomics: exploring the world around biological systems. *Expert Review of Proteomics* 9, 561-575.
- Arnold, R.J., and Reilly, J.P. (1998). Fingerprint matching of E-coli strains with matrix-assisted laser desorption ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. *Rapid Communications in Mass Spectrometry* 12, 630-636.
- Barber, M., Bordoli, R.S., Sedgwick, R.D., and Tyler, A.N. (1981). Fast atom bombardment of solids as an ion-source in mass spectrometry. *Nature* 293, 270-275.
- Blondiaux, N., Gaillot, O., and Courcol, R.J. (2010). MALDI-TOF mass spectrometry to identify clinical bacterial isolates: Evaluation in a teaching hospital. *Pathologie Biologie* 58, 55-57.
- Borio, L., Inglesby, T., Peters, C.J., Schmaljohn, A.L., Hughes, J.M., Jahrling, P.B., Ksiazek, T., Johnson, K.M., Meyerhoff, A., O'Toole, T., *et al.* (2002). Hemorrhagic fever viruses as biological weapons - Medical and public health management. *Jama-Journal of the American Medical Association* 287, 2391-2405.
- Bosio, C.M. (2011). The subversion of the immune system by *Francisella tularensis*. *Frontiers in microbiology* 2.
- Brauer, H.A., Lampe, P.D., Yasui, Y.Y., Hamajima, N., and Stolowitz, M.L. (2010). Biochips that sequentially capture and focus antigens for immunoaffinity MALDI-TOF MS: A new tool for biomarker verification. *Proteomics* 10, 3922-3927.
- Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007). Isotope-labeled protein standards. *Molecular & Cellular Proteomics* 6, 2139-2149.

- Bulut, O.C., Dyckhoff, G., Splettstoesser, W., Nemeth, J., Klauschen, F., Penzel, R., Plinkert, P.K., Simon, C., Weichert, W., and Stenzinger, A. (2013). Unmasked: When a Clinically Malignant Disease Turns Out Infectious. A Rare Case of Tularemia. *International Journal of Surgical Pathology* 21, 76-81.
- Carbonnelle, E., Beretti, J.-L., Cottyn, S., Quesne, G., Berche, P., Nassif, X., and Ferroni, A. (2007). Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization - Time of flight mass spectrometry. *Journal of Clinical Microbiology* 45, 2156-2161.
- Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A. (2004). The need for guidelines in publication of peptide and protein identification data - Working group on publication guidelines for peptide and protein identification data. *Molecular & Cellular Proteomics* 3, 531-533.
- Casadevall, A., and Pirofski, L.A. (2000). Host-pathogen interactions: Basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and Immunity* 68, 6511-6518.
- Casadevall, A., and Pirofski, L.A. (2001). Host-pathogen interactions: The attributes of virulence. *Journal of Infectious Diseases* 184, 337-344.
- Cassagne, C., Ranque, S., Normand, A.C., Fourquet, P., Thiebault, S., Planard, C., Hendrickx, M., and Piarroux, R. (2011). Mould Routine Identification in the Clinical Laboratory by Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry. *Plos One* 6, 9.
- Chen, W.-J., Tsai, P.-J., and Chen, Y.-C. (2008). Functional Nanoparticle-Based Proteomic Strategies for Characterization of Pathogenic Bacteria. *Analytical Chemistry* 80, 9612-9621.
- Chenau, J., Fenaille, F., Caro, V., Haustant, M., Diancourt, L., Klee, S.R., Junot, C., Ezan, E., Goossens, P.L., and Becher, F. (2014a). Identification and Validation of Specific Markers of *Bacillus anthracis* Spores by Proteomics and Genomics Approaches. *Molecular & Cellular Proteomics* 13, 716-732.
- Chenau, J., Fenaille, F., Ezan, E., Morel, N., Lamourette, P., Goossens, P.L., and Becher, F. (2011). Sensitive Detection of *Bacillus anthracis* Spores by Immunocapture and Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Chemistry* 83, 8675-8682.
- Chenau, J., Fenaille, F., Simon, S., Filali, S., Volland, H., Junot, C., Carniel, E., and Becher, F. (2014b). Detection of *Yersinia pestis* in Environmental and Food Samples by Intact Cell Immunocapture and Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Chemistry* 86, 6144-6152.
- Christie-Oleza, J.A., Miotello, G., and Armengaud, J. (2013). Proteogenomic Definition of Biomarkers for the Large *Roseobacter* Clade and Application for a Quick Screening of New Environmental Isolates. *Journal of Proteome Research* 12, 5331-5339.
- Christie-Oleza, J.A., Pina-Villalonga, J.M., Guerin, P., Miotello, G., Bosch, R., Nogales, B., and Armengaud, J. (2012). Shotgun nanoLC-MS/MS proteogenomics to document MALDI-TOF biomarkers for screening new members of the *Ruegeria* genus. *Environ Microbiol.*
- Christopher, G.W., Cieslak, T.J., Pavlin, J.A., and Eitzen, E.M. (1997). Biological warfare - A historical perspective. *Jama-Journal of the American Medical Association* 278, 412-417.
- Clark, A.E., Kaleta, E.J., Arora, A., and Wolk, D.M. (2013). Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: a Fundamental Shift in the Routine Practice of Clinical Microbiology. *Clinical Microbiology Reviews* 26, 547-603.

- Collectif (1970). Health aspects of chemical and biological weapons (Geneva).
- Collectif (1978). Ebola haemorrhagic fever in Sudan, 1976. Report of a WHO/International Study Team. *56*, 247-270.
- Conlan, J.W. (2011). Tularemia vaccines: recent developments and remaining hurdles. *Future Microbiology* *6*, 391-405.
- de Godoy, L.M.F., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Froehlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* *455*, 1251-U1260.
- Demirev, P.A., Ho, Y.P., Ryzhov, V., and Fenselau, C. (1999). Microorganism identification by mass spectrometry and protein database searches. *Analytical Chemistry* *71*, 2732-2738.
- Demirev, P.A., Lin, J.S., Pineda, F.J., and Fenselau, C. (2001). Bioinformatics and mass spectrometry for microorganism identification: Proteome-wide post-translational modifications and database search algorithms for characterization of intact *H-pylori*. *Analytical Chemistry* *73*, 4566-4573.
- Dennis, D.T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Layton, M., *et al.* (2001). Tularemia as a biological weapon - Medical and public health management. *Jama-Journal of the American Medical Association* *285*, 2763-2773.
- Derbes, V.J. (1966). Demussis and great plague of 1348 - a forgotten episode of bacteriological warfare. *Journal of the American Medical Association* *196*, 59-&.
- Desvaux, M., Hebraud, M., Talon, R., and Henderson, I.R. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends in Microbiology* *17*, 139-145.
- Dupieris, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., and Bruley, C. (2009). A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* *25*.
- Durighello, E., Christie-Oleza, J.A., and Armengaud, J. (2014). Assessing the Exoproteome of Marine Bacteria, Lesson from a RTX-Toxin Abundantly Secreted by *Phaeobacter* Strain DSM 17395. *Plos One* *9*, 10.
- Dworzanski, J.P., Deshpande, S.V., Chen, R., Jabbour, R.E., Snyder, A.P., Wick, C.H., and Li, L. (2006). Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *Journal of Proteome Research* *5*, 76-87.
- Dworzanski, J.P., Dickinson, D.N., Deshpande, S.V., Snyder, A.P., and Eckenrode, B.A. (2010). Discrimination and Phylogenomic Classification of *Bacillus anthracis-cereus-thuringiensis* Strains Based on LC-MS/MS Analysis of Whole Cell Protein Digests. *Analytical Chemistry* *82*, 145-155.
- Dworzanski, J.P., Snyder, A.P., Chen, R., Zhang, H.Y., Wishart, D., and Li, L. (2004). Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring. *Analytical Chemistry* *76*, 2355-2366.
- Ecker, D.J., Sampath, R., Massire, C., Blyn, L.B., Hall, T.A., Eshoo, M.W., and Hofstadler, S.A. (2008). Innovation - Ibis T5000: a universal biosensor approach for microbiology. *Nature Reviews Microbiology* *6*, 553-558.



- Elias, J.E., and Gygi, S.P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in molecular biology* (Clifton, NJ) **604**, 55-71.
- Emonet, S., Shah, H.N., Cherkaoui, A., and Schrenzel, J. (2010). Application and use of various mass spectrometry methods in clinical microbiology. *Clinical Microbiology and Infection* **16**, 1604-1613.
- Fagerquist, C.K., Garbus, B.R., Miller, W.G., Williams, K.E., Yee, E., Bates, A.H., Boyle, S., Harden, L.A., Cooley, M.B., and Mandrell, R.E. (2010). Rapid Identification of Protein Biomarkers of *Escherichia coli* O157:H7 by Matrix-Assisted Laser Desorption Ionization-Time-of-Flight-Time-of-Flight Mass Spectrometry and Top-Down Proteomics. *Analytical Chemistry* **82**, 2717-2725.
- Farlow, J., Smith, K.L., Wong, J., Abrams, M., Lytle, M., and Keim, P. (2001). *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *Journal of Clinical Microbiology* **39**, 3186-3192.
- Feldmann, H., Jones, S., Klenk, H.D., and Schnittler, H.J. (2003). Ebola virus: from discovery to vaccine. *Nature Reviews Immunology* **3**, 677-685.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science* **246**, 64-71.
- Fenner, F., Henderson, D.A., Arita, I., Ježek, Z., and Ladnyi, I.D. (1988). Smallpox and its eradication, Vol 6, World Health Organization Geneva edn (World Health Organization Geneva).
- Fey, P.D., Dempsey, M.M.P., Olson, M.E., Chrustowski, M.S., Engle, J.L., Jay, J.J., Dobson, M.E., Kalasinsky, K.S., Shea, A.A., Wen, P.C., *et al.* (2007). Molecular analysis of *Francisella tularensis* subspecies *tularensis* and *holarctica*. *American Journal of Clinical Pathology* **128**, 926-935.
- Foley, J.E., and Nieto, N.C. (2010). Tularemia. *Veterinary Microbiology* **140**, 332-338.
- Forsman, M., Sandstrom, G., and Jaurin, B. (1990). Identification of *Francisella* species and discrimination of type-A and type-B strains of *Francisella-tularensis* by 16S ribosomal-RNA analysis. *Applied and Environmental Microbiology* **56**, 949-955.
- Fournier, P.-E., Drancourt, M., Colson, P., Rolain, J.-M., La Scola, B., and Raoult, D. (2013). Modern clinical microbiology: new challenges and solutions. *Nature Reviews Microbiology* **11**, 574-585.
- Francis, E. (1925). Tularemia. *JAMA: The Journal of the American Medical Association* **84**, 1243-1250.
- Francis, E., Mayne, B., and Lake, G.C. (1922). Tularemia Francis 1921: A new disease of man. *Hyg Lab Bull* **130**, 1-87.
- Frank, A.M. (2009a). A Ranking-Based Scoring Function for Peptide-Spectrum Matches. *Journal of Proteome Research* **8**, 2241-2252.
- Frank, A.M. (2009b). Predicting Intensity Ranks of Peptide Fragment Ions. *Journal of Proteome Research* **8**, 2226-2240.
- Garcia Del Blanco, N., Dobson, M.E., Vela, A.I., De La Puente, V.A., Gutierrez, C.B., Hadfield, T.L., Kuhnert, P., Frey, J., Dominguez, L., and Rodriguez Ferri, E.F. (2002). Genotyping of *Francisella tularensis* strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing. *Journal of clinical microbiology* **40**, 2964-2972.

- Gillet, L.C., Navarro, P., Tate, S., Roest, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* 11.
- Gohlke, R.S. (1959). Time-of-flight mass spectrometry and gas-liquid partition chromatography. *Analytical Chemistry* 31, 535-541.
- Gopal, J., Narayana, J.L., and Wu, H.-F. (2011). TiO<sub>2</sub> nanoparticle assisted mass spectrometry as biosensor of *Staphylococcus aureus*, key pathogen in nosocomial infections from air, skin surface and human nasal passage. *Biosensors & Bioelectronics* 27, 201-206.
- Guo, Z., Liu, Y., Li, S., and Yang, Z. (2009). Interaction of bacteria and ion-exchange particles and its potential in separation for matrix-assisted laser desorption/ionization mass spectrometric identification of bacteria in water. *Rapid Communications in Mass Spectrometry* 23, 3983-3993.
- Gurycova, D. (1998). First isolation of *Francisella tularensis* subsp. *tularensis* in Europe. *European Journal of Epidemiology* 14, 797-802.
- Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17, 994-999.
- Haag, A.M., Taylor, S.N., Johnston, K.H., and Cole, R.B. (1998). Rapid identification and speciation of *Haemophilus* bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 33, 750-756.
- Hepburn, M.J., and Simpson, A.J.H. (2008). Tularemia: current diagnosis and treatment options. *Expert Review of Anti-Infective Therapy* 6, 231-240.
- Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991). Matrix-assisted laser desorption ionization mass-spectrometry of biopolymers. *Analytical Chemistry* 63, A1193-A1202.
- Ho, Y.-P., and Reddy, P.M. (2010). Identification of Pathogens by Mass Spectrometry. *Clinical Chemistry* 56, 525-536.
- Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz, J., and Giordani, A.B. (1981). Sequence-analysis of polypeptides by collision activated dissociation on a triple quadrupole mass-spectrometer. *Biomedical Mass Spectrometry* 8, 397-408.
- Jabbour, R.E., Deshpande, S.V., Wade, M.M., Stanford, M.F., Wick, C.H., Zulich, A.W., Skowronski, E.W., and Snyder, A.P. (2010). Double-Blind Characterization of Non-Genome-Sequenced Bacteria by Mass Spectrometry-Based Proteomics. *Applied and Environmental Microbiology* 76, 3637-3644.
- Jarman, K.H., Cebula, S.T., Saenz, A.J., Petersen, C.E., Valentine, N.B., Kingsley, M.T., and Wahl, K.L. (2000). An algorithm for automated bacterial identification using matrix-assisted laser desorption/ionization mass spectrometry. *Analytical Chemistry* 72, 1217-1223.
- Jarman, K.H., Daly, D.S., Petersen, C.E., Saenz, A.J., Valentine, N.B., and Wahl, K.L. (1999). Extracting and visualizing matrix-assisted laser desorption/ionization time-of-flight mass spectral fingerprints. *Rapid Communications in Mass Spectrometry* 13, 1586-1594.

- Jernigan, J.A., Stephens, D.S., Ashford, D.A., Omenaca, C., Topiel, M.S., Galbraith, M., Tapper, M., Fisk, T.L., Zaki, S., Popovic, T., *et al.* (2001). Bioterrorism-related inhalational anthrax: The first 10 cases reported in the United States. *Emerging Infectious Diseases* 7, 933-944.
- Jiang, J., Parker, C.E., Fuller, J.R., Kawula, T.H., and Borchers, C.H. (2007). An immunoaffinity tandem mass spectrometry (iMALDI) assay for detection of *Francisella tularensis*. *Analytica Chimica Acta* 605, 70-79.
- Johansson, A., Forsman, M., and Sjostedt, A. (2004). The development of tools for diagnosis of tularemia and typing of *Francisella tularensis*. *Apmis* 112, 898-907.
- Johansson, A., and Petersen, J.M. (2010). Genotyping of *Francisella tularensis*, the Causative Agent of Tularemia. *Journal of Aoac International* 93, 1930-1943.
- Johansson, A., Tomaso, H., Padeshki, P., Sjöstedt, a., Silman, N., and Pilo, P. (2012). *Francisella tularensis*: Tularemia. In *BSL3 and BSL4 Agents Epidemiology, Microbiology and Practical Guidelines*, M. Elschner, S. Cutler, M. Weidmann, and P. Butaye, eds. (WILEY-BLACKWELL), pp. 71-84.
- Johnson, E.M., Ellis, W.R., Jr., Powers, L.S., and Wysocki, V.H. (2009). Affinity Capture Mass Spectrometry of Biomarker Proteins Using Peptide Ligands from Biopanning. *Analytical Chemistry* 81, 5999-6005.
- Johnson, R.S., Martin, S.A., and Biemann, K. (1988). Collision-induced fragmentation of (M+H)<sup>+</sup> ions of peptides - side-chain specific sequence ions. *International Journal of Mass Spectrometry and Ion Processes* 86, 137-154.
- Johansson, B., Lofas, S., and Lindquist, G. (1991). Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface-plasmon resonance sensors. *Analytical Biochemistry* 198, 268-277.
- Jones, C.L., Napier, B.A., Sampson, T.R., Llewellyn, A.C., Schroeder, M.R., and Weiss, D.S. (2012). Subversion of Host Recognition and Defense Systems by *Francisella* spp. *Microbiology and Molecular Biology Reviews* 76, 383-404.
- Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. *Analytical Chemistry* 60, 2299-2301.
- Keim, P., Johansson, A., and Wagner, D.M. (2007). Molecular epidemiology, evolution, and ecology of *Francisella*. *Francisella Tularensis: Biology, Pathogenicity, Epidemiology, and Biodefense* 1105, 30-66.
- Kim, Y.E., Yi, S.Y., Lee, C.-S., Jung, Y., and Chung, B.H. (2012). Gold patterned biochips for on-chip immuno-MALDI-TOF MS: SPR imaging coupled multi-protein MS analysis. *Analyst* 137, 386-392.
- Klemm, R., Schattschneider, S., Jahn, T., Hlawatsch, N., Julich, S., Becker, H., and Gartner, C. (2013). Development of a versatile lab-on-a-chip enzyme assay platform for pathogen detection in CBRNE scenarios. Paper presented at: Conference on Micro- and Nanotechnology Sensors, Systems, and Applications V (Baltimore, MD: Spie-Int Soc Optical Engineering).
- Kliem, M., and Sauer, S. (2012). The essence on mass spectrometry based microbial diagnostics. *Current Opinion in Microbiology* 15, 397-402.
- Koehling, H.L., Bittner, A., Mueller, K.-D., Buer, J., Becker, M., Ruebben, H., Rettenmeier, A.W., and Mosel, F. (2012). Direct identification of bacteria in urine samples by matrix-assisted laser

desorption/ionization time-of-flight mass spectrometry and relevance of defensins as interfering factors. *Journal of Medical Microbiology* 61, 339-344.

Krishnamurthy, T., and Ross, P.L. (1996). Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. *Rapid Communications in Mass Spectrometry* 10, 1992-1996.

Krishnamurthy, T., Ross, P.L., and Rajamani, U. (1996). Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 10, 883-888.

Kull, S., Pauly, D., Stoermann, B., Kirchner, S., Staemmler, M., Dorner, M.B., Lasch, P., Naumann, D., and Dorner, B.G. (2010). Multiplex Detection of Microbial and Plant Toxins by Immunoaffinity Enrichment and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Analytical Chemistry* 82, 2916-2924.

Lacroix, C., Gicquel, A., Sendid, B., Meyer, J., Accoceberry, I., Francois, N., Morio, F., Desoubes, G., Chandenier, J., Kauffmann-Lacroix, C., *et al.* (2014). Evaluation of two matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) systems for the identification of *Candida* species. *Clinical Microbiology and Infection* 20, 153-158.

Lasch, P., Nattermann, H., Erhard, M., Staemmler, M., Grunow, R., Bannert, N., Appel, B., and Naumann, D. (2008). MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Analytical Chemistry* 80.

Lavigne, J.P., Espinal, P., Dunyach-Remy, C., Messad, N., Pantel, A., and Sotto, A. (2013). Mass spectrometry: a revolution in clinical microbiology? *Clinical Chemistry and Laboratory Medicine* 51, 257-270.

Lay, J.O., Jr. (2001). MALDI-TOF mass spectrometry of bacteria. In *Mass Spectrom Rev* (United States: 2002 John Wiley & Sons, Inc.), pp. 172-194.

Li, S., Guo, Z., Liu, Y., Yang, Z., and Hui, H.K. (2009). Integration of microfiltration and anion-exchange nano particles-based magnetic separation with MALDI mass spectrometry for bacterial analysis. *Talanta* 80, 313-320.

Li, S., Guo, Z., Wu, H.-F., Liu, Y., Yang, Z., and Woo, C.H. (2010). Rapid analysis of Gram-positive bacteria in water via membrane filtration coupled with nanoprobes-based MALDI-MS. *Analytical and Bioanalytical Chemistry* 397, 2465-2476.

Li, Y., Powell, D.A., Shaffer, S.A., Rasko, D.A., Pelletier, M.R., Leszyk, J.D., Scott, A.J., Masoudi, A., Goodlett, D.R., Wang, X., *et al.* (2012). LPS remodeling is an evolved survival strategy for bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, 8716-8721.

Liang, X.L., Lubman, D.M., Rossi, D.T., Nordblom, G.D., and Barksdale, C.M. (1998). On probe immunoaffinity extraction by matrix-assisted laser desorption/ionization mass spectrometry. *Analytical Chemistry* 70, 498-503.

Lin, Y.S., Tsai, P.J., Weng, M.F., and Chen, Y.C. (2005). Affinity capture using vancomycin-bound magnetic nanoparticles for the MALDI-MS analysis of bacteria. *Analytical Chemistry* 77, 1753-1760.

MacCoss, M.J., Wu, C.C., and Yates, J.R., 3rd (2002). Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* 74, 5593-5599.

Madonna, A.J., Basile, F., Furlong, E., and Voorhees, K.J. (2001). Detection of bacteria from biological mixtures using immunomagnetic separation combined with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 15, 1068-1074.

Madonna, A.J., Van Cuyk, S., and Voorhees, K.J. (2003). Detection of *Escherichia coli* using immunomagnetic separation and bacteriophage amplification coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 17, 257-263.

Massey, E.D., and Mangiafi, J.A. (1974). Microagglutination test for detecting and measuring serum agglutinins of *Francisella-tularensis*. *Applied Microbiology* 27, 25-27.

Massey, E.D., and Mangiafi, J.A. (1974). MICROAGGLUTINATION TEST FOR DETECTING AND MEASURING SERUM AGGLUTININS OF *FRANCISELLA-TULARENSIS*. *Applied Microbiology* 27, 25-27.

McCoy, G.W. (1911). A plague-like disease in rodents. *Public Health Bulletin* 43, 53-71.

Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., and Hunt, D.F. (2006). The utility of ETD mass spectrometry in proteomic analysis. *Biochimica Et Biophysica Acta-Proteins and Proteomics* 1764, 1811-1822.

Moore, Z.S., Seward, J.F., and Lane, J.M. (2006). Smallpox. *Lancet* 367, 425-435.

Mueller, L.N., Brusniak, M.-Y., Mani, D.R., and Aebersold, R. (2008). An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *Journal of Proteome Research* 7, 51-61.

Munson, M.S.B., and Field, F.H. (1966). Chemical ionization mass spectrometry .I. General introduction. *Journal of the American Chemical Society* 88, 2621-&.

Mwenye, K.S., Siziya, S., and Peterson, D. (1996). Factors associated with human anthrax outbreak in the Chikupo and Ngandu villages of Murewa district in Mashonaland East Province, Zimbabwe. *The Central African journal of medicine* 42, 312-315.

Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Molecular & Cellular Proteomics* 11.

Needham, B.D., and Trent, M.S. (2013). Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews Microbiology* 11, 467-481.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data - The protein inference problem. *Molecular & Cellular Proteomics* 4, 1419-1440.

Nesvizhskii, A.I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* 4, 787-797.

Niyogi, S.K. (2005). Shigellosis. *Journal of Microbiology* 43, 133-143.

Ochoa, M.L., and Harrington, P.B. (2005). Immunomagnetic isolation of enterohemorrhagic *Escherichia coli* O157 : H7 from ground beef and identification by matrix-assisted laser

desorption/ionization time-of-flight mass spectrometry and database searches. *Analytical Chemistry* 77, 5258-5267.

Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* 4, 709-712.

Olsen, J.V., and Mann, M. (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* 101, 13417-13422.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* 1, 376-386.

Paizs, B., and Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 24, 508-548.

Patrie, S.M., and Mrksich, M. (2007). Self-assembled monolayers for MALDI-TOF mass Spectrometry for Immunoassays of human protein antigens. *Analytical Chemistry* 79, 5878-5887.

Pavkova, I., Reichelova, M., Larsson, P., Hubalek, M., Vackova, J., Forsberg, A., and Stulik, J. (2006). Comparative proteome analysis of fractions enriched for membrane-associated proteins from *Francisella tularensis* subsp *tularensis* and *F. tularensis* subsp *holarctica* strains. *Journal of Proteome Research* 5, 3125-3134.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.

Perry, R.D., and Fetherston, J.D. (1997). *Yersinia pestis* - Etiologic agent of plague. *Clinical Microbiology Reviews* 10, 35-&.

Picard, G., Lebert, D., Louwagie, M., Adrait, A., Huillet, C., Vandenesch, F., Bruley, C., Garin, J., Jaquinod, M., and Brun, V. (2012). PSAQ (TM) standards for accurate MS-based quantification of proteins: from the concept to biomedical applications. *Journal of Mass Spectrometry* 47, 1353-1363.

Pineda, F.J., Lin, J.S., Fenselau, C., and Demirev, P.A. (2000). Testing the significance of microorganism identification by mass spectrometry and proteome database search. *Analytical Chemistry* 72, 3739-3744.

Pirofski, L.-a., and Casadevall, A. (2012). Q&A: What is a pathogen? A question that begs the point. *Bmc Biology* 10.

Raetz, C.R.H., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry* 71, 635-700.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11, 601.

Ross, P.L., Huang, Y.L.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., *et al.* (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* 3, 1154-1169.

Rusconi, F. (2011). Manuel de spectrométrie de masse à l'usage des biochimistes (Lavoisier).



- Rychert, J., Burnham, C.-A.D., Bythrow, M., Garner, O.B., Ginocchio, C.C., Jennemann, R., Lewinski, M.A., Manji, R., Mochon, A.B., Procop, G.W., *et al.* (2013). Multicenter Evaluation of the Vitek MS Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry System for Identification of Gram-Positive Aerobic Bacteria. *Journal of Clinical Microbiology* 51, 2225-2231.
- Ryhage, R. (1964). USE OF MASS SPECTROMETER AS DETECTOR + ANALYZER FOR EFFLUENTS EMERGING FROM HIGH TEMPERATURE GAS LIQUID CHROMATOGRAPHY COLUMNS. *Analytical Chemistry* 36, 759-8.
- Sack, D.A., Sack, R.B., Nair, G.B., and Siddique, A.K. (2004). Cholera. *Lancet* 363, 223-233.
- Saenz, A.J., Petersen, C.E., Valentine, N.B., Gantt, S.L., Jarman, K.H., Kingsley, M.T., and Wahl, K.L. (1999). Reproducibility of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for replicate bacterial culture analysis. *Rapid Communications in Mass Spectrometry* 13, 1580-1585.
- Sandrin, T.R., Goldstein, J.E., and Schumaker, S. (2013). MALDI TOF MS profiling of bacteria at the strain level: A review. *Mass Spectrometry Reviews* 32, 188-217.
- Santic, M., Molmeret, M., Klose, K.E., Jones, S., and Kwak, Y.A. (2005). The *Francisella tularensis* pathogenicity island protein IgIC and its regulator MglA are essential for modulating phagosome biogenesis and subsequent bacterial escape into the cytoplasm. *Cellular Microbiology* 7, 969-979.
- Sato, T., Fujita, H., Ohara, Y., and Homma, M. (1990). Microagglutination test for early and specific serodiagnosis of tularemia. *Journal of Clinical Microbiology* 28, 2372-2374.
- Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M., and Geider, K. (2008). Classification and Identification of Bacteria by Mass Spectrometry and Computational Analysis. *Plos One* 3, 10.
- Sauer, S., and Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology* 8, 74-82.
- Schlosser, G., Kacer, P., Kuzma, M., Szilagyi, Z., Sorrentino, A., Manzo, C., Pizzano, R., Malorni, L., and Pocsfalvi, G. (2007). Coupling immunomagnetic separation on magnetic beads with matrix-assisted laser desorption ionization-time of flight mass spectrometry for detection of staphylococcal enterotoxin B. *Applied and Environmental Microbiology* 73, 6945-6952.
- Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., and Splettstoesser, W. (2010). Identification of *Francisella tularensis* by Whole-Cell Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: Fast, Reliable, Robust, and Cost-Effective Differentiation on Species and Subspecies Levels. *Journal of Clinical Microbiology* 48, 1061-1069.
- Spencer, R.C. (2003). *Bacillus anthracis*. *Journal of Clinical Pathology* 56, 182-187.
- Splettstoesser, W., Guglielmo-Viret, V., Seibold, E., and Thullier, P. (2010). Evaluation of an Immunochromatographic Test for Rapid and Reliable Serodiagnosis of Human Tularemia and Detection of *Francisella tularensis*-Specific Antibodies in Sera from Different Mammalian Species. *Journal of Clinical Microbiology* 48, 1629-1634.
- Staples, J.E., Kubota, K.A., Chalcraft, L.G., Mead, P.S., and Petersen, J.M. (2006). Epidemiologic and molecular analysis of human tularemia, United States, 1964-2004. *Emerging Infectious Diseases* 12, 1113-1118.

- Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews Molecular Cell Biology* 5, 699-711.
- Stevenson, L.G., Drake, S.K., and Murray, P.R. (2010). Rapid Identification of Bacteria in Positive Blood Culture Broths by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *Journal of Clinical Microbiology* 48, 444-447.
- Stolowitz, M., L., and Stephan, A., H. (2009). SOLID-PHASE AFFINITY-BASED METHOD FOR PREPARING AND MANIPULATING AN ANALYTE-CONTAINING SOLUTION.
- Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., *et al.* (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959-964.
- Svensson, K., Larsson, P., Johansson, D., Bystrom, M., Forsman, M., and Johansson, A. (2005). Evolution of subspecies of *Francisella tularensis*. *Journal of Bacteriology* 187, 3903-3908.
- Szalo, I.M., Taminiau, B., and Mainil, J. (2006). Lipopolysaccharide of *Escherichia coli*: structure, biosynthesis and roles. *Annales De Medecine Veterinaire* 150, 108-124.
- Tabb, D.L., Saraf, A., and Yates, J.R. (2003). GutenTag: High-throughput sequence tagging via an empirically derived fragmentation model. *Analytical Chemistry* 75, 6415-6421.
- Taylor, C.F., Binz, P.-A., Aebersold, R., Affolter, M., Barkovich, R., Deutsch, E.W., Horn, D.M., Huehmer, A., Kussmann, M., Lilley, K., *et al.* (2008). Guidelines for reporting the use of mass spectrometry in proteomics. *Nature Biotechnology* 26, 860-861.
- Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.-A., Julian, R.K., Jr., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., *et al.* (2007). The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnology* 25, 887-893.
- Taylor, J.A., and Johnson, R.S. (1997). Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* 11, 1067-1075.
- Thirumalapura, N.R., Morton, R.J., Ramachandran, A., and Malayer, J.R. (2005). Lipopolysaccharide microarrays for the detection of antibodies. *Journal of Immunological Methods* 298, 73-81.
- Torok, T.J., Tauxe, R.V., Wise, R.P., Livengood, J.R., Sokolow, R., Mauvais, S., Birkness, K.A., Skeels, M.R., Horan, J.M., and Foster, L.R. (1997). A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *Jama-Journal of the American Medical Association* 278, 389-395.
- Touelle, M., Mirabella, B., Guerin, P., de la Tour, C.B., Boissard, S., Hong Ha, N., Blanchard, L., Servant, P., de Groot, A., Sommer, S., *et al.* (2012). A comparative proteomic approach to better define *Deinococcus nucleoid* specificities. *Journal of Proteomics* 75.
- van Belkum, A., Welker, M., Erhard, M., and Chatellier, S. (2012). Biomedical Mass Spectrometry in Today's and Tomorrow's Clinical Microbiology Laboratories. *Journal of Clinical Microbiology* 50, 1513-1517.
- Vogler, A.J., Birdsell, D., Price, L.B., Bowers, J.R., Beckstrom-Sternberg, S.M., Auerbach, R.K., Beckstrom-Sternberg, J.S., Johansson, A., Clare, A., Buchhagen, J.L., *et al.* (2009a). Phylogeography of *Francisella tularensis*: Global Expansion of a Highly Fit Clone. *Journal of Bacteriology* 191, 2474-2484.



Vogler, A.J., Birdsell, D., Wagner, D.M., and Keim, P. (2009b). An optimized, multiplexed multi-locus variable-number tandem repeat analysis system for genotyping *Francisella tularensis*. *Letters in Applied Microbiology* 48, 140-144.

Wu, H.-F., Gopal, J., and Manikandan, M. (2012). Future perspective of nanoparticle interaction-assisted laser desorption/ionization mass spectrometry for rapid, simple, direct and sensitive detection of microorganisms. *Journal of Mass Spectrometry* 47, 355-363.

Yates, J.R., III (2011). A century of mass spectrometry: from atoms to proteomes. *Nature Methods* 8, 633-637.

Yoder, J.D., Chen, T.S., Gagnier, C.R., Vemulapalli, S., Maier, C.S., and Hruby, D.E. (2006). Pox proteomics: mass spectrometry analysis and identification of *Vaccinia* virion proteins. *Virology Journal* 3.

Zubarev, R.A., Kelleher, N.L., and McLafferty, F.W. (1998). Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 120, 3265-3266.

# Annexes



# Annexes

## 1 Table des annexes

**Annexe 1** : Valeurs des 100  $m/z$  les plus intenses sur les spectres MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS

**Annexe 2** : Fonctions des protéines identifiées par nanoLC-MS/MS dans le protéome de *Francisella tularensis* subsp. *holarctica* LVS inactivée au TFA.

**Annexe 3** : Méthode de traitement des spectres sur FlexAnalysis. Programmé en langage BASIC

**Annexe 4** : Algorithme donnant un score aux spectres selon la probabilité de présence d'un organisme dans un échantillon. Programmé en langage Octave.

## Annexe 1: Valeurs des 100 $m/z$ les plus intenses sur les spectres MALDI-TOF de *Francisella tularensis* subsp. *holarctica* LVS

Les 100 plus intenses $m/z$ pour chaque spectre				Liste des 44 $m/z$ (moyens) présents dans au moins 3 spectres sur 4
Spectrum 1	Spectrum 2	Spectrum 3	Spectrum 4	
3048.4	2950.5	2962.3	2982.3	3364.4
3049.4	2952.2	2973.8	2983.2	3365.8
3050.1	2962.9	2974.2	3000.0	3369.3
3156.7	2970.0	3013.9	3245.4	3613.5
3173.8	3117.0	3014.6	3364.7	3616.4
3364.7	3152.0	3015.5	3365.7	3619.1
3365.6	3240.9	3033.8	3366.9	3897.1
3369.7	3242.8	3043.4	3369.1	4015.4
3489.2	3244.8	3149.0	3387.6	4113.3
3520.1	3248.7	3151.1	3419.5	4121.4
3547.4	3359.5	3164.6	3894.7	4734.8
3591.4	3363.7	3178.7	3894.7	5120.0
3613.8	3365.5	3244.0	3998.0	5132.8
3616.2	3388.2	3265.1	4015.0	5136.8
3619.7	3389.8	3265.2	4114.6	5141.9
3662.9	3612.1	3360.8	4118.4	5431.9
3751.6	3613.8	3361.5	4121.9	5734.8
3898.2	3617.8	3366.4	4123.4	6179.0
4016.4	3619.1	3367.8	4279.4	6359.3
4017.5	3899.9	3368.9	4734.9	6741.2
4120.7	3957.1	3384.3	4836.8	7142.1
4123.0	4013.4	3385.0	4836.8	7386.4
4124.0	4014.9	3455.1	4838.8	7463.6
4736.4	4111.7	3562.1	5115.5	7802.7
4839.4	4113.2	3612.8	5116.5	8132.8
4841.4	4116.9	3612.8	5133.6	8249.0
4842.5	4278.8	3615.1	5138.0	8268.5
5120.2	4280.8	3618.3	5142.3	8270.1
5128.4	4283.4	3725.9	5432.4	8406.1
5131.6	4734.1	3895.8	5734.4	8614.6
5134.1	5118.0	3995.3	6171.9	8810.1
5136.1	5119.2	3997.9	6174.0	9073.5
5139.1	5120.3	4061.2	6179.9	9475.5
5142.9	5126.7	4063.8	6282.1	9682.6
5146.7	5133.4	4112.2	6360.4	9698.7
5435.9	5136.5	4121.7	6733.7	10242.5
5733.7	5141.3	4397.7	6741.3	10273.9
5735.5	5146.4	4399.7	7138.1	10370.0
5737.7	5148.2	4585.7	7142.0	10871.0
6102.7	5431.3	4733.7	7386.6	11478.0
6176.7	5734.6	4744.3	7390.4	12360.5
6177.8	6177.9	4835.9	7464.2	12718.5
6179.3	6285.4	5119.5	7546.0	16809.9

6287.9	6359.4	5132.7	7546.0	16952.3
6358.2	6740.9	5132.8	7803.2	
6737.0	7142.5	5136.7	8122.8	
6741.3	7385.7	5141.3	8133.1	
7141.8	7456.9	5180.8	8249.6	
7143.5	7463.2	5429.7	8268.4	
7382.6	7545.8	5429.7	8269.8	
7383.8	7549.7	5431.0	8406.2	
7385.2	7802.0	5432.0	8549.0	
7386.9	8042.2	5731.4	8566.3	
7463.3	8132.6	5734.5	8772.5	
7465.5	8247.9	6172.9	8790.1	
7467.6	8256.9	6178.9	8810.0	
7803.3	8268.2	6183.6	9073.4	
8125.1	8404.9	6352.9	9448.5	
8133.2	8476.0	6356.6	9475.5	
8249.3	8477.8	6738.9	9682.7	
8268.8	8565.7	6741.1	9697.9	
8270.8	8614.5	7140.3	9698.7	
8406.7	8772.5	7142.2	10242.2	
8414.6	8809.6	7386.2	10252.8	
8614.5	8906.6	7386.2	10256.6	
8810.0	9061.2	7388.6	10260.2	
8813.3	9073.1	7389.5	10273.5	
8906.3	9429.7	7463.6	10288.2	
8906.4	9433.8	7802.4	10292.1	
9074.1	9437.6	8132.5	10358.7	
9459.7	9441.3	8249.2	10370.0	
9475.9	9474.8	8268.8	10390.1	
9674.5	9668.2	8269.7	10773.2	
9683.2	9670.1	8406.5	10774.9	
9699.7	9681.6	8476.5	10870.1	
10243.6	9697.7	8611.2	11006.5	
10275.2	9720.5	8611.7	11006.5	
10292.7	10240.9	8614.7	11477.5	
10368.0	10272.9	8810.6	12201.7	
10371.5	10360.3	9073.3	12203.8	
10373.5	10369.1	9475.8	12203.8	
10706.2	10773.2	9682.9	12311.4	
10872.6	10855.4	9699.3	12338.3	
11009.3	10869.8	10243.3	12360.6	
11480.2	11005.8	10274.3	12387.7	
12206.1	11075.9	10369.5	12565.4	
12315.8	11477.1	10871.5	12578.3	
12319.3	11560.3	11479.3	12580.8	
12360.9	12304.6	11479.5	12718.1	
12381.4	12313.3	12203.8	12727.6	
12383.7	12359.1	12315.9	12728.7	
12722.3	12566.3	12361.3	16793.1	
12929.7	12717.5	12376.4	16809.3	
13061.9	13054.1	12379.7	16944.4	
13062.6	16622.2	12719.9	16950.4	

---

13064.5	16630.6	13057.3	16951.5
16643.2	16808.1	16812.3	16952.5
16646.7	16945.4	16952.3	16956.9
16821.9	16952.3	17225.9	17229.1
17241.3	17223.2	17230.4	17233.8

---

## Annexe 2 : Fonctions des protéines identifiées par nanoLC-MS/MS dans le protéome de *Francisella tularensis* subsp. *holarctica* LVS inactivée au TFA.

---

### Protéines Ribosomales

---

emb|CAJ78695.1| 50S ribosomal protein L30  
 emb|CAJ78694.1| 30S ribosomal protein S5  
 emb|CAJ79464.1| 30S ribosomal protein S18  
 emb|CAJ80177.1| 30S ribosomal protein S16  
 emb|CAJ78691.1| 30S ribosomal protein S8  
 emb|CAJ78684.1| 50S ribosomal protein L16  
 emb|CAJ78689.1| 50S ribosomal protein L5  
 emb|CAJ78688.1| 50S ribosomal protein L24  
 emb|CAJ78699.1| 30S ribosomal protein S13  
 emb|CAJ78686.1| 30S ribosomal protein S17  
 emb|CAJ78693.1| 50S ribosomal protein L18  
 emb|CAJ78674.1| 30S ribosomal protein S7  
 emb|CAJ78700.1| 30S ribosomal protein S11  
 emb|CAJ79977.1| 30S ribosomal protein S15  
 emb|CAJ78679.1| 50S ribosomal protein L23  
 emb|CAJ78962.1| 50S ribosomal protein L28  
 emb|CAJ79844.1| 50S ribosomal protein L35  
 emb|CAJ79626.1| 50S ribosomal subunit protein L13  
 emb|CAJ79625.1| 30S ribosomal protein S9  
 emb|CAJ79843.1| 50S ribosomal protein L20  
 emb|CAJ80184.1| 50S ribosomal protein L7/L12  
 emb|CAJ78676.1| 30S ribosomal protein S10  
 emb|CAJ80187.1| 50S ribosomal protein L11  
 emb|CAJ80186.1| 50S ribosomal protein L1  
 emb|CAJ78665.1| 30S ribosomal protein S2  
 emb|CAJ78511.1| 30S ribosomal protein S20  
 emb|CAJ78690.1| 30S ribosomal protein S14  
 emb|CAJ78677.1| 50S ribosomal protein L3  
 emb|CAJ78687.1| 50S ribosomal protein L14  
 emb|CAJ78683.1| 30S ribosomal protein S3  
 emb|CAJ79891.1| 50S ribosomal protein L27  
 emb|CAJ78682.1| 50S ribosomal protein L22  
 emb|CAJ79486.1| 30S ribosomal protein S21  
 emb|CAJ78685.1| 50S ribosomal protein L29  
 emb|CAJ79465.1| 50S ribosomal protein L9  
 emb|CAJ79742.1| 50S ribosomal protein L31  
 emb|CAJ78703.1| 50S ribosomal protein L17

---

### Protéines ayant un lien avec le ribosome

---

emb|CAJ80190.1| elongation factor Tu (EF-Tu)



emb|CAJ78675.1| elongation factor G (EF-G)  
 emb|CAJ79618.1| Sigma-54 modulation protein  
 emb|CAJ78666.1| protein chain elongation factor EF-Ts

---

#### Promoteur de l'assemblage de l'ARN polymerase

emb|CAJ79972.1| DNA-directed RNA polymerase omega subunit

---

#### Protéines liées à l'ADN

emb|CAJ79334.1| Histone-like protein HU form B

---

#### Protéines chaperonnes

emb|CAJ80153.1| Chaperone protein, groEL  
 sp|P94797.1|CH10\_FRATH 10 kDa chaperonin; GroES protein; Protein Cpn10  
 emb|CAG45902.1| Chaperone protein dnaK (heat shock protein family 70 protein)  
 emb|CAJ79337.1| host factor I for bacteriophage Q beta replication

---

#### Protéines de transport

emb|CAJ79286.1| preprotein translocase family protein  
 emb|CAJ79577.1| acyl carrier protein

---

#### Protéines membranaires associant la synthèse de l'ATP et/ou le transport de proton

emb|CAJ80236.1| ATP synthase alpha chain  
 emb|CAJ80235.1| ATP synthase gamma chain

---

#### Repliement des protéines

emb|CAJ79832.1| Peptidyl-prolyl cis-trans isomerase

---

#### Protéines membranaires

sp|P18149.1|MP17\_FRATH 17 kDa major membrane protein; TUL4  
 emb|CAJ79767.1| outer membrane associated protein  
 emb|CAJ78776.1| Peptidoglycan-associated lipoprotein

---

#### Protéines présentes dans la mitochondrie, métabolisme carbone/ nitrogen

emb|CAJ78710.1| NAD(P)-specific glutamate dehydrogenase

---

#### Enzymes de régulation de la voie de synthèse des acides gras

emb|CAJ80031.1| Acetyl-CoA carboxylase, biotin carboxyl carrier protein subunit  
 emb|CAJ79748.1| Acetyl-CoA carboxylase beta subunit  
 emb|CAJ78736.1| Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha  
 emb|CAJ79576.1| 3-oxoacyl-[acyl-carrier-protein] synthase II

---

#### Enzyme catalysant le succinyl-CoA en succinate

emb|CAJ79992.1| Succinyl-CoA synthetase beta chain

---

#### Enzyme participant au cycle de l'acide citrique

emb|CAJ79028.1| isocitrate dehydrogenase

---

Enzyme catalisant la synthèse du citrate

emb|CAJ80228.1| citrate synthase

---

Enzyme réticulant les chaînes de peptidoglycanes pour construire les murs cellulaires

emb|CAJ79499.1| D-alanyl-D-alanine carboxypeptidase (Penicillin binding protein) family protein

---

Enzyme catalisant la conversion de la dethiobiotine (DTB) en biotine

emb|CAJ79711.1| biotin synthase

---

Protéine associée à la glutaredoxine

emb|CAJ80231.1| Glutaredoxin-related protein

---

Protéine de l'activité oxydoreductase

emb|CAJ79454.1| AhpC/TSA family protein

emb|CAJ79057.1| conserved hypothetical protein

---

Croissance intracellulaire

emb|CAJ78552.1| intracellular growth locus, subunit A

---

Protéine impliquée dans la division cellulaire

sp|Q9ZAW3.1|FTSZ\_FRATH Cell division protein FtsZ

---

Lipoprotéine

emb|CAJ79085.1| hypothetical lipoprotein]

---

## Annexe 3 : Méthode de traitement des spectres sur FlexAnalysis. Programmé en langage BASIC

```

Option Explicit
Sub Main
UndoAllProcessing()
' ReplaceCalibration()
' PrintOut "fff"

Dim spectrumName$
spectrumName = Split(Spectra(1).Name)(0)
'LSLin

Dim FoldersNames$

Dim arrayId() As String
arrayId = Split(Path, "\")
Dim arrayLen As Integer
arrayLen = UBound( arrayId )

'Dim i As Integer
'For i = 0 To UBound( arrayId )
'    MsgBox( arrayId(i) )
'    MsgBox( Format(UBound( arrayId )) ) )
'Next i

'MsgBox(Len(spectrumName))

Dim spectreFilename As String
spectreFilename = Replace(Name, "\", "~")

Dim outputFilename$
outputFilename = "C:\Documents and Settings\ed230673\Bureau\FlexAnalysisBATCH\"+
spectreFilename + ".txt"

Dim spectre As Spectrum

Dim dRange(1) As Double
dRange(0) = 0.0
dRange(1) = 20000.0

ResultSpectra(1).Process(faSubtractBaseline)
ResultSpectra(1).Process(faSmooth)
ResultSpectra(1).Process(faFindPeaks)

ResultSpectra(1).Enabled = True

'Spectra(1).Process(faSmooth)
'Spectra(1).Process(faPeakFind)
'faBaselineSubtract = 1

'Save
'Process
'Processes the Spectrum. The processed Spectrum Is added To the Spectra collection
of
'this spectrum.
'Process(Algorithm, [Range])
'Parameter:
'Algorithm FAProcessingAlgorithm.
'Range Double[] Optional. The Range specifies the processing range As Min. x, Max.
x-
'value Or As Min. x, Max. x, Min. y, Max. y values.

```

```

'Return value: The processed Spectrum.

'FAProcessingAlgorithm
'faSmooth
'Smooth With Method parameters.
'faBaselineSubtract
'Subtract baseline With method parameters.
'faPeakFind
'Find peaks With method parameters.

'Export
'Exports the Spectrum.
'Syntax:
'Export (Filename, precisionX, precisionY)
'Parameter:
'Filename String. Filename of the export Spectrum. The file extension of
'the filename determines the export file Format.
'precisionX Long. Number of Decimal digits of the x-values (mass)
'precisionY Long. Number of Decimal digits of the y-values (intensity)

'ResultSpectra(1).Export(outputFilename, 2, 0)

Dim sComment1 As String
sComment1 = ResultSpectra(1).Parameters.Find("CMT1")

Dim sComment2 As String
sComment2 = ResultSpectra(1).Parameters.Find("CMT2")


Dim oFSO As Object
Set oFSO = CreateObject("Scripting.FileSystemObject")

Dim oTextStream As Object
Set oTextStream = oFSO.OpenTextFile(outputFilename, 8, True)
oTextStream.WriteLine(Path)
'oTextStream.WriteLine(sComment1)
'oTextStream.WriteLine(sComment2)
oTextStream.WriteLine(";;;;;;;;;;")
oTextStream.WriteLine("m/z" & ";" & "Int.")

Dim myMassList As MassList
Set myMassList = ResultSpectra(1).MassList

Dim oPeak As MassPeakInfo
Dim sMass As String
Dim sInt As String
For Each oPeak In myMassList
    sMass = Format(oPeak.MassToChargeValue)
    sInt = Format(oPeak.Intensity)
    oTextStream.WriteLine(sMass & ";" & sInt)
Next

oTextStream.WriteLine(vbCrLf)
oTextStream.Close

End Sub

```

## Annexe 4 : Algorithme donnant un score aux spectres selon la probabilité de présence d'un organisme dans un échantillon. Programmé en langage Octave.

```
#recherche de la présence des biomarqueurs
#A partir du fichier avec les masses il faut dire la fourchette de masse dans
laquelle il faut se placer pour voir s'il y a un pic.
#Liste des masses première colonne (0) deuxième ligne(1))

% warning('off', 'Octave:possible-matlab-short-circuit-operator');
% warning('off','all');
% do_braindead_shortcircuit_evaluation (1)

arg_list = argv ();
filename = arg_list{1};
# filename2 = arg_list{2};

% input=arg_list{1};
% printf ("%s", input);

A=dlmread ("biomarqueursFT.csv", ",", 1,0);

#longueur de la liste de masse
length(A);

#creation de ma matrice des max et des min à 300 ppm près
for i=1:length (A)
    Max(i)=A(i)+A(i)*0.0003;
    Min(i)=A(i)-A(i)*0.0003;
endfor
Max;
Min;

#creation du vecteur des pics presents Mc
Mc=zeros(length(A),1);
#chargement de la matrice des pics de l'échantillon
#lecture à partir de la colonne 0 et de la ligne 3

Ech=dlmread(filename,"",3,0);
for i=1:length(A);
    for j=1:length(Ech);
        test=Min(i)<=Ech(j)&Ech(j)<=Max(i);
        if test;
            Mc(i)=1;
        endif
    endfor
endfor
Mc;

#import du vecteur des probas P
# il est dans le fichier "biomarqueursFT.csv", 3ème colonne (2) 2ème ligne (1)
P=dlmread("BiomarqueursFT.csv",";",1,2);
P=P';

#Creation du vecteurs M
length(Mc);
Mc=Mc';
#M matrice des pics absents
M=ones(1,length(Mc))-Mc;
#P est le vecteur des probabilités
PMC=P.*Mc;
```

```

PM=ones(1,length(P))-P.*M;
#produit des proba pour les pics présents
pmc=1;
pm=1;
for i=1:length(P)
    if PMc(1,i)~=0
        pmc=pmc*PMc(1,i);
    endif
    if PM(1,i)~=1
        pm=pm*PM(1,i);
    endif
endfor
pmc;
pm;

#score d'indentification
#on fixe le seuil du nombre de biomarqueurs nécessaires pour pouvoir procéder à
l'indentification(correspond à 20% des pics présents)
S=0.2;
#si il y a moins de 20 % des biomarqueurs présents alors da est nul
if sum(Mc)<S*length(Mc)
    da=0;
else
    da=1-(pmc*(1-pm));
endif
#affichage du nombre de pics
NbPics=sum(Mc);
#étape de réponse: selon la valeur de da dire si la bactérie est présente ou non

if da>=0.7
    result='FT is highly likely to be present';
endif

test = da>=0.15 & da<0.7;
if test
    result='FT is likely to be present';
endif

test = da>=0.05 & da<0.15;
if test
    result='inconclusive';
endif

test = da>=0.01 & da<0.05;
if test
    result='FT is unlikely to be present';
endif
if da<0.01
    result='FT is highly unlikely to be present';
endif
printf ("%s\tda\t%f\tNbPics\t%d\t%s\t\n",filename,da,NbPics,result);

```



## 2 Participation à l'écriture de revue et d'articles

### 2.1 Articles

Le premier article est le fruit du travail abordé au cours de mon stage de Mastère spécialisé en Bio-ingénierie effectué au LBSP entre février et octobre 2011. Les résultats ont été valorisés par cet article rédigé et argumenté durant ma thèse. Il s'agit d'un travail de séquençage *de novo* sur une toxine de *Phaeobacter gallaeciensis*, une bactérie marine de la clade des *Roseobacters*. Malgré une présence abondante de cette protéine, elle est récalcitrante à l'identification classique par spectrométrie de masse en tandem. Nous proposons une stratégie basée sur l'utilisation de plusieurs enzymes protéolytiques pour une meilleure analyse des exoprotéomes, constitués de protéines en général récalcitrantes aux protéases.

Les données supplémentaires sont disponibles à l'adresse suivante :

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0089691#s4>

---

**Assessing the exoproteome of marine bacteria, lesson from a RTX-toxin abundantly secreted by *Phaeobacter* strain DSM 17395.**

E. Durighello, J.A. Christie-Oleza, J. Armengaud

PLoS ONE 9(2): e89691. doi:10.1371/journal.pone.0089691

p. 187 à 196 du mémoire de thèse

---

Le second article présente la première partie des résultats détaillés dans ce manuscrit. Il s'agit de l'étude protéogénomique de *Francisella tularensis* subsp. *holarctica* LVS afin de définir un ensemble de protéines permettant de différencier les différentes espèces et sous-espèces des bactéries du genre des *Francisella* et ainsi de conclure sur leur virulence à partir de spectres MALDI-TOF. Cet article a été accepté par le journal *Analytical Chemistry*

Les données supplémentaires sont disponibles à l'adresse suivante :

<http://pubs.acs.org/doi/suppl/10.1021/ac501840g>

---

**Proteogenomic biomarkers for identification of *Francisella* species and subspecies by MALDI-TOF mass spectrometry**

E. Durighello, L. Bellanger, E. Ezan, J. Armengaud

Accepté le 12 septembre 2014 dans *Analytical Chemistry*

p. 197 à 201 du mémoire de thèse

---



## 2.2 Revue

Suite à l'écriture de l'article sur la RTX-toxine de la bactérie marine *Phaeobacter gallaeciensis* et la présentation de résultats sur l'identification de bactéries par spectrométrie de masse au congrès CIESM 2013 à Marseille, j'ai participé à l'écriture d'une revue invitée qui détaille en quoi la protéomique peut être un atout en biologie marine.

---

**Proteomics meets blue biotechnology: A wealth of novelties and opportunities.**

E.M. Hartmann, E. Durighello, O. Pible, B. Nogales, F. Beltrametti, R. Bosch, J.A. Christie-Oleza, J. Armengaud

Marine genomics (2014) <http://dx.doi.org/10.1016/j.margen.2014.04.003>

p. 203 à 211 du mémoire de thèse

---

# Assessing the Exoproteome of Marine Bacteria, Lesson from a RTX-Toxin Abundantly Secreted by *Phaeobacter* Strain DSM 17395

Emie Durighello, Joseph Alexander Christie-Oleza, Jean Armengaud\*

CFA, DSV, IRFR, Lab Biochim System Perturb, Ragnols-sur-Cèze, France

## Abstract

Bacteria from the *Roseobacter* clade are abundant in surface marine ecosystems as over 10% of bacterial cells in the open ocean and 20% in coastal waters belong to this group. In order to document how these marine bacteria interact with their environment, we analyzed the exoproteome of *Phaeobacter* strain DSM 17395. We grew the strain in marine medium, collected the exoproteome and catalogued its content with high-throughput nanoLC-MS/MS shotgun proteomics. The major component represented 60% of the total protein content but was refractory to either classical proteomic identification or proteogenomics. We *de novo* sequenced this abundant protein with high-resolution tandem mass spectra which turned out being the 53 kDa RTX-toxin ZP\_02147451. It comprised a peptidase M10 serralyisin domain. We explained its recalcitrance to trypsin proteolysis and proteomic identification by its unusual low number of basic residues. We found this is a conserved trait in RTX-toxins from *Roseobacter* strains which probably explains their persistence in the harsh conditions around bacteria. Comprehensive analysis of exoproteomes from environmental bacteria should take into account this proteolytic recalcitrance.

**Citation:** Durighello E, Christie-Oleza JA, Armengaud J (2014) Assessing the Exoproteome of Marine Bacteria, Lesson from a RTX-Toxin Abundantly Secreted by *Phaeobacter* Strain DSM 17395. PLoS ONE 9(2): e89691. doi:10.1371/journal.pone.0089691

**Editor:** Daniel Gillet, CEA (Atomic and Alternative Energies Commission), France

**Received:** December 4, 2013; **Accepted:** January 21, 2014; **Published:** February 24, 2014

**Copyright:** © 2014 Durighello et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** JA Christie-Oleza was supported by a fellowship from the Fundación Ramón Areces. The authors thank the Commissariat à l'Energie Atomique et aux Energies Alternatives for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: jean.armengaud@cea.fr

## Introduction

The extremely high diversity of marine microorganisms was shown ever since the first large metagenomic shotgun analysis carried out on total DNA extracted from filtered seawater samples collected from the Sargasso Sea near Bermuda [1]. Heterotrophic bacterioplankton communities in near-surface marine pelagic environments mostly comprise *Alphaproteobacteria* and *Gammaproteobacteria*, as well as *Sphingobacteria* and *Flavobacteria* (both belonging to the *Bacteroidetes* superphylum). As shown recently with environmental metagenomic data, prominent components belong to the SAR11, *Roseobacter* and SAR86 clades together with the OM60/NOR5 cluster [2]. The *Roseobacter* lineage is a phylogenetically coherent group of *Alphaproteobacteria* comprising up to 20% of marine microbial communities, especially in coastal and polar waters [3]. However, the *Roseobacter* clade is a physiologically heterogeneous group showing a high variety of lithoheterotrophic strategies. Despite a free-living lifestyle, *Roseobacter* strains are also found as planktonic or larger pluricellular symbionts [4].

Today, in-depth analysis of a microbial proteome relies on the use of high-resolution hybrid tandem mass spectrometers coupled to high-pressure liquid chromatography systems able to resolve complex peptide mixtures [5,6]. A comprehensive list of proteins and their quantitation in several physiological conditions by such shotgun approaches are a rather straightforward method [7,8]. Exoproteomes comprise the fraction of proteins secreted by cellular systems and proteins arising from cell lysis [9]. These are

generally subjected to intense proteolytic activities and most components have a short half-life. Nevertheless, key components involved in nutrient imports, cell motility, signaling and interactions with other organisms should be refractory to proteolysis in order to be efficient. Analysis of bacterial secretomes is an important issue for the discovery of novel bioactive compounds and to determine how organisms interact within their environment. We have shown that microbial exoproteomes can be comprehensively analyzed by high throughput shotgun proteomics [10,11]. The use of the most recent generation of hybrid tandem mass spectrometers revealed a high diversity of protein virulence factors in the pathogenic *Bacillus cereus* bacteria [11]. To date, only a few exoproteome analyses have been reported on environmental avirulent isolates [10,12]. In this case, the data obtained helped elucidating the ecological distinctness among strains of the *Roseobacter* clade. Microbial exoproteomes are also insightful for understanding the molecular mechanisms of bacterial colonization of biotic or abiotic surfaces, i.e. bacterial adhesion and biofilm formation [13]. Their exploration gains ground with the current development of proteomic tools for environmental microbiology [14].

*Phaeobacter gallaeciensis* BS107 (previously *Roseobacter gallaeciensis* BS107) is the type species of the *Phaeobacter* genus [15]. This strain was obtained from larval cultures and collectors of the scallop *Pecten maximus* [16]. Other species of the genus have since been described: *Phaeobacter inhibens*, isolated from the German Wadden

Sea [15], *Phaeobacter daeponensis*, isolated from a tidal flat of the Yellow Sea at Daepo Beach in Korea [17], and *Phaeobacter arcticus*, a psychrophilic species isolated from marine sediments of the Arctic Ocean [18]. *Phaeobacter inhibens* was in-depth characterized with proteomics and metabolomics approaches highlighting its versatile metabolism [19,20,21,22]. The type species *Phaeobacter strain DSM 17395* was shown to be able to produce a new tropolone derivative, tropodithietic acid, which exhibits strong antibiotic properties against marine bacteria of various taxa and marine algae [23,24]. Additional studies led to the identification of nine new tropenoid compounds belonging to the bacterial roseobacticide family [25]. More recently, it has been shown that *Phaeobacter strain DSM 17395* switches its secreted molecule metabolism to the production of potent and selective algacides in response to p-coumaric acid, an algal lignin breakdown product that is symptomatic of aging algae. This switch converts *Phaeobacter strain DSM 17395* into an opportunistic pathogen of its algal host [26] and highlights the interest of the study of its secretome. The type strain *Phaeobacter gallaeciensis BS107T* has been deposited at public culture collections worldwide. Recently, it was shown that the *Phaeobacter strain DSM 17395*, previously annotated as *Phaeobacter gallaeciensis BS107T*, exhibits a much closer affiliation to *Phaeobacter inhibens* DSM 16374T and should thus be allocated to this species [27].

With the intention to further characterize the ecological distinctness of *Roseobacters* members and how they interact with their environment, we analyzed the secreted proteins of *Phaeobacter strain DSM 17395*. For this, we grew the strain in marine medium, collected the exoproteome and catalogued its protein content with high throughput nanoLC-MS/MS shotgun proteomics. Interestingly, the major protein in this exoproteome was recalcitrant to classical proteomic identification. Indeed it appears very abundant on SDS-PAGE migration but no protein was matching this band after proteomic analysis. A *de novo* sequencing strategy identified peptides of a RTX-like toxin which led to re-analyze proteomics data with no-enzyme specificity. This RTX-like toxin was then clearly identified. We found that this protein and its homologues exhibit unusually low content in basic residues. From these results, we proposed a general strategy to obtain more comprehensive exoproteome analyses of environmental bacteria.

## Materials and Methods

### *P. gallaeciensis* Growth and Exoproteome Preparation

Four flasks containing 40 mL of marine broth (Difco) were inoculated with *Phaeobacter strain DSM 17395* cells previously grown on marine agar plates. The cultures were incubated at 30°C under 180 rpm agitation until mid-exponential phase growth ( $OD_{600} = 0.6$ ). Cultures were centrifuged at 3,000 *g* for 10 min at 20°C and supernatants were carefully filtered through two low protein-binding filters of 0.45  $\mu$ m (Millex-Hv) and then 0.22  $\mu$ m (Millex-GV) diameter pore (Millipore) in order to eliminate bacterial cells. Supernatants were frozen prior protein precipitation. Proteins from the supernatant were precipitated with trichloroacetic acid as previously described [10]. The resulting protein pellets were dissolved into 90  $\mu$ l of lithium dodecyl sulfate- $\beta$ -mercaptoethanol protein gel sample buffer (Invitrogen), incubated at 99°C for 5 min and briefly centrifuged prior SDS-PAGE.

### SDS-PAGE and Trypsin-*in gel* Proteolysis

A volume of 20  $\mu$ l of the concentrated exoproteome of *Phaeobacter strain DSM 17395* was deposited per well onto 10% Bis-Tris NuPAGE gels (Invitrogen). SDS-PAGE was carried out using 1X 3-(N-morpholino) propanesulfonic acid solution (Invitro-

gen) in a XCell SureLock Mini-cell (Invitrogen) under a constant voltage of 200 V. At first, a long migration was carried out in order to fully resolve the exoproteome according to its molecular weight. For the proteomic shotgun analysis, the entire exoproteome was allowed to enter the polyacrylamide gel in a short migration (3 mm). Both gels were stained with SimplyBlue SafeStain, a ready-to-use Coomassie G-250 stain (Invitrogen). SeeBlue Plus2 (Invitrogen) was used as a molecular weight marker. Densitometry analysis of the long-migrated SDS-PAGE gel was carried out with the molecular imager GS-800 calibrated densitometer (BioRad) and the Quantity One software (version 4.6.9, BioRad) in order to establish the relative abundance of the main protein band. Densitometric relative ratio was obtained by comparing the density signal of each band to the total signal of the whole gel lane. These measurements were done in triplicate. Polyacrylamide gel bands (equivalent in volume to 50  $\mu$ l) were cut and processed for *in-gel* proteolysis with Trypsin Sequencing Grade (Roche) followed by the ProteaseMax protocol (Promega) as previously described [11].

### NanoLC-MS/MS Analysis

Peptide digests were resolved on an UltiMate 3000 LC system (Dionex-LC Packings) as previously described [28] prior to MS/MS measurements done with a LTQ-Orbitrap XL (ThermoFisher). For the FT/IT procedure, parameters used for the tandem mass spectrometry measurements were those previously described [29]. During the FT/FT procedure, the LC gradient used was a 90 min gradient with aqueous solvent A (0.1% HCOOH) and solvent B (0.1% HCOOH/80% CH<sub>3</sub>CN) developed as follows: 5–60% B in 90 min, 60–90% B in 1 min, 90% B during 10 min, 90–5% B in 1 min and 5% B during 19 min. The “full scan” mass spectrum was measured from *m/z* 300 to 1800 *th* and the charge states considered were +2 and +3. A scan cycle was started with a high resolution full scan (resolution = 30,000) with the Orbitrap analyzer, followed by a high resolution MS/MS scan of the secondary ions obtained after fragmentation, again with the Orbitrap analyzer but with a slightly lower resolution (resolution = 15,000). A 30 sec dynamic exclusion was set. The activation type used was CID with a standard normalized collision energy set at 35.

### MS/MS Database MASCOT Search and Protein Quantification by Spectral Count

Peak lists were generated with the MASCOT DAEMON software (v2.3.2, Matrix Science) from the LC-MS/MS raw data using default parameters. The MASCOT 2.3.02 search engine (Matrix Science) was used to search all MS/MS spectra against a database corresponding to a complete list of annotated Coding Domain Sequences (CDS) from *Phaeobacter strain DSM 17395* (NCBI RefSeq: NC\_018290.1, NC\_018291.1, NC\_018288.1, and NC\_018287.1). Bovine trypsin and the 22 most common keratin contaminants were included as a follow-up. Searches for peptides were first performed with the following parameters: mass tolerance of 5 ppm on the parent ion and 0.5 Da on the MS/MS (FT/IT procedure), static modifications of carboxamidomethylated Cys (+57.0215), and dynamic modification of oxidized Met (+15.9949). High-resolution MS/MS (FT/FT procedure) searches were performed with mass tolerance of 0.1 Da on the MS/MS. To process the data resulting from trypsin proteolysis, the maximum number of missed cleavage was set at 2. All peptide matches with a score above its peptidic identity threshold (set at  $p < 0.01$ ) with the ORF database and rank 1 were filtered with the IRMa 1.28.0 software [30]. A protein was only validated when at least two peptides had been assigned. In a second stage, we performed



additional searches for both “semi-trypsin” and “no-enzyme” protease specificities. Protein abundance was evaluated in the shotgun analysis by MS/MS spectral counts as established by Liu and co-workers [31] and previously described [32]. Normalized Spectral count Abundance Factors (NSAF) was calculated as defined by Paoletti et al. [33].

### De novo Assignment of MS/MS Spectra

Peak lists were submitted to Peaks Studio 5.3 software (Bioinformatics Solutions Inc). At first, a data refinement step was performed with a quality threshold set at 0.65. *De novo* sequencing was then carried out with the following parameters: a parent mass error tolerance of 5.0 ppm, a fragment mass error tolerance of 0.1 Da, cysteine carbamidomethylation (+57.02) as fixed modification, and methionine oxidation (+15.99) as variable modification. More stringent criteria were applied with a parent mass error tolerance of 2 ppm with high resolution mode for both MS and MS/MS. Successively, trypsin, semi-trypsin and no-enzyme were chosen as enzyme specificities. The list of assigned spectra obtained with MASCOT using *Phaeobacter strain DSM 17395* CDS database was compared with Peaks Studio results for each samples. Precursor mass spectra annotated with the Peaks Studio software but not interpreted with the Mascot engine gave candidates for manual annotation. Results were then sorted by their Average of Local Confidence (ALC) in order to choose the best spectra to annotate.

## Results and Discussion

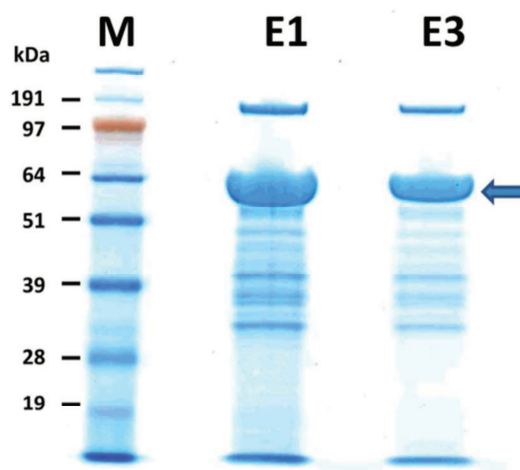
### The most Abundant Protein within the Exoproteome of *Phaeobacter strain DSM 17395* is Recalcitrant to Standard Proteomic Identification

The resulting exoproteome of the marine bacterium *Phaeobacter strain DSM 17395* when grown in Marine Broth was collected and resolved by SDS-PAGE. Remarkably, an abundant protein band migrating at a molecular weight of about 55 kDa was observed in the protein profile. This 55 kDa band accounts for 60 ( $\pm 2$ ) % of the total exoproteome as estimated by densitometry (**Figure 1**). With the objective of identifying most of the components of this exoproteome, we analyzed the sample with the previously developed shotgun procedure [10]. The whole exoproteome was excised from a short SDS-PAGE migration (3 mm) as a single small polyacrylamide band. After trypsinization, the peptide mixture was injected into a C18 reversed phase chromatography column coupled to a high resolution LTQ-Orbitrap XL hybrid mass spectrometer. The acquired tandem mass spectra were assigned with the Mascot software using as database the complete list of annotated CDS from *Phaeobacter strain DSM 17395*. The dataset comprised 2321 MS/MS spectra, from which 1112 could be assigned to specific peptides from *Phaeobacter strain DSM 17395* (**Table S1 in File S1**). These spectra corresponded to 478 different peptides which validated the presence of 75 proteins with at least two different peptides (**Table S2 in File S1**). In terms of spectral counts, the first 15 proteins accounted for half of the total number of assigned spectra. The most detected proteins were an outer membrane porin (ZP\_02147263), a flagellin-like protein (COG1344, ZP\_02145485), a hemolysin-type calcium-binding protein (ZP\_02144235), and the periplasmic phosphate-binding subunit of a phosphate ABC transporter (ZP\_02143869) with 86, 55, 52, and 41 MS/MS spectra, respectively (**Table S2 in File S1**). Despite all these proteins were expected to be secreted, none was abundant enough (8% of total MS/MS spectra) as to be considered the highly secreted protein seen in **Figure 1**. Furthermore, their molecular weights (32, 28, 116, and 37 kDa,

respectively) differ from that expected for this protein ( $\approx 55$  kDa). Therefore, the shotgun results were in disagreement with the SDS-PAGE information and the abundant 55 kDa protein could not be identified. To solve this issue, we fully resolved the exoproteome onto a SDS-PAGE gel as shown in **Figure 1** and analyzed by nanoLC-MS/MS the protein content of the 55 kDa polyacrylamide band. Although only one major polypeptide was expected to be present in this polyacrylamide band, we evidenced 48 proteins identified with at least 2 peptides. **Table 1** reports the ten first proteins as ranked by their Mascot score (assigned MS/MS spectra and corresponding peptide characteristics are reported in **Table S3 in File S1**). Surprisingly we observed the protein GroEL Chaperonin (ZP\_02145205) as the most detected protein with 61 MS/MS spectra although this protein is not expected to be secreted by bacteria. Another unexpected result was the presence in this polyacrylamide band of the 116 kDa protein ZP\_02144235. Its tryptic sequence coverage showed that both the N- and C-termini were confirmed with proteomic-detected peptides, but the internal sequence was poorly covered (**Table S3 in File S1**). We presumed that this protein may be subjected to specific maturation such as a proteolytic cleavage. Again, none of the proteins shown in **Table 1** were in sufficient abundance as to be considered the abundant protein seen in **Figure 1**. Whether the expected abundant 55 kDa protein was encoded on the genome but missed during the annotation stage was also investigated by a new Mascot search using a six-frame translation of the whole nucleic acid sequence, as previously described [34]. This proteogenomic approach remained unsuccessful with no considerable findings (**Data not shown**). Therefore, the abundantly secreted protein of *Phaeobacter strain DSM 17395* remained unidentified with standard proteomic identification.

### De novo Sequencing with High Resolution Data Revealed the Identity of the Abundantly Secreted Protein

We carried out a search to manually assign spectra that had not been automatically assigned to a tryptic peptide comprised in the



**Figure 1. SDS-PAGE of the exoproteome of *Phaeobacter strain DSM 17395*.** Exoproteins were resolved by a long migration on a 10% SDS-PAGE and stained with SimplyBlue SafeStain (Invitrogen). Lane **M**: SeeBlue Plus2 molecular weight range marker (Invitrogen). Lane **E1**: *Phaeobacter strain DSM 17395* exoproteome grown in Marine Broth (20  $\mu$ g). Lane **E2**: *Phaeobacter strain DSM 17395* exoproteome grown in Marine Broth (8  $\mu$ g). The 55 kDa major component is indicated with an arrow.

doi:10.1371/journal.pone.0089691.g001

**Table 1.** List of the first ten proteins identified from the 55 kDa exoproteome band of *P. gallaeciensis* DSM<sup>1</sup>.

Identification rank	Accession	Functional annotation	Number of residues	Protein Mascot Score	Molecular weight (Da)	Sequence Coverage (%)	Number of unique peptide	Spectral Count
1	ZP_02145205	Chaperonin GroEL	550	1632	57749	62	28	61
2	ZP_02146523	Oligopeptide/dipeptide ABC transporter*	518	929	57342	32	15	37
3	ZP_02144235	Hemolysin-type calcium-binding protein	1112	915	115931	15	13	35
4	ZP_02145689	Extracellular solute-binding protein, family 1	577	884	64572	36	15	33
5	ZP_02143510	Ser/Thr protein phosphatase/nucleotidase, putative	525	594	56013	36	12	25
6	ZP_02146184	Extracellular solute-binding protein, family 5	566	559	61051	31	11	21
7	ZP_02145455	Flagellar hook-associated protein	482	553	50979	28	9	16
8	ZP_02147263	Outer membrane porin	313	508	31831	28	9	21
9	ZP_02144479	Extracellular solute-binding protein, family 5	528	413	57617	17	7	20
10	ZP_02144175	Acetate hydratase	895	431	96861	13	10	19

<sup>1</sup>Detected with at least three different peptides.

\*Periplasmic component.

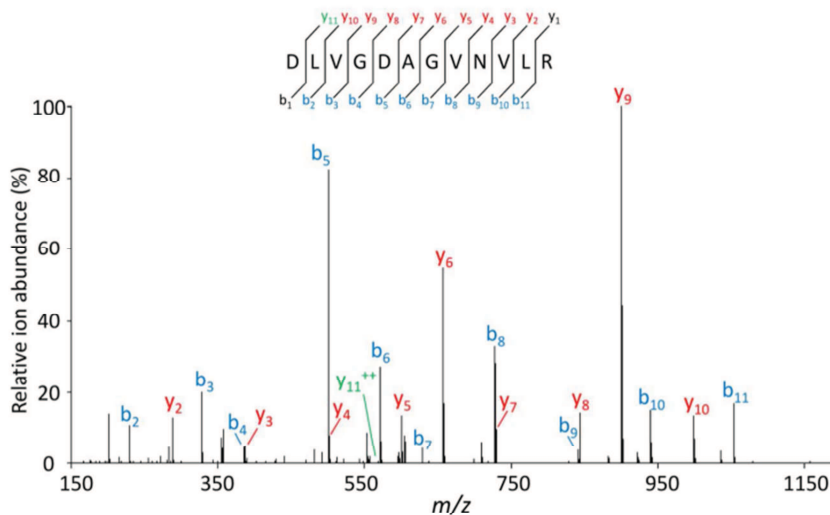
doi:10.1371/journal.pone.0089691.t001

*Phaeobacter* strain DSM 17395 CDS database. For this *de novo* sequencing strategy, we reanalyzed the peptides from the 55 kDa protein band by nanoLC-MS/MS with a FT/FT procedure which enables getting higher resolution MS/MS spectra and better mass accuracy (15,000 in the orbitrap analyzer instead of the 3,000 previously done with the LTQ analyzer). From this new dataset acquisition, 246 MS/MS spectra were assigned directly with the Mascot software using the complete list of annotated CDSs from *Phaeobacter* strain DSM 17395 as a database. The same dataset was also searched with the *de novo* sequencing software Peaks Studio 5.3 identifying 258 new MS/MS spectra that had not been assigned by Mascot. We manually annotated these spectra checking all peptide sequence possibilities. **Figure 2** shows the annotation of a representative MS/MS spectrum with a  $m/z$  ratio measured at 614.33826 amu corresponding to a di-charged peptide with a semi-tryptic sequence, DLVGDAQVNVLR. Peak assignment was confident with the detection of 10 monoprotonated *b* ions and 9 monoprotonated *y* ions, except that the second and eleventh residues could be either leucine or isoleucine. We could also assign the first residue with the detection of the  $y_{11}$  diprotonated fragment ion. A total of five sequences could be confidently assigned because of MS/MS mass accuracy. The Blast analysis of each of these five tentative peptides revealed strict identities to the ZP\_02147451 protein from *Phaeobacter* strain DSM 17395. Curiously, all five *de novo* sequenced peptides corresponded to semi-tryptic peptides. With these data in hands, we finally supposed that the abundant exoprotein was ZP\_02147451, annotated as a hemolysin-type calcium binding protein (RTX-toxin). Its theoretical molecular weight, 52624 Da, matched the expected molecular weight estimated from the SDS-PAGE gel.

### High Resolution Data Allowed a High Coverage of ZP\_02147451 with Non-tryptic Peptides

To confirm the identity of the protein and because *de novo* sequencing of some MS/MS spectra had shown the presence of abundant non-tryptic peptides, we re-analyzed the FT/FT experimental data with the Mascot software using “no-enzyme” as the enzyme parameter. **Table S4 in File S1** lists the resulting 228 MS/MS spectra assigned to 128 distinct peptides belonging to proteins of the *Phaeobacter* strain DSM 17395 bacterium. As expected, less MS/MS spectra are assigned with the no-enzyme search mode than with the trypsin search mode, due to the loss of discrimination introduced by the larger search space. This time the hemolysin-type calcium binding protein ZP\_02147451 was clearly the most abundant protein identified accounting 67 MS/MS spectra assigned (30% of the spectral counts). A total of 36 distinct peptides were found for this protein and allowed 51% coverage of the whole sequence (**Figure 3**). According to this “no-enzyme” Mascot search, 24 peptides were semi-tryptic, 10 items do not arise from a specific cleavage, and only 1 is a tryptic peptide despite the trypsinization protocol applied. Peptides were evenly distributed along the protein sequence except the peptidase unit region spanning residues 240 to 279 (**Figure 3**). **Table 2** shows the comparative list of proteins assigned in the two Mascot search modes (“trypsin” or “no-enzyme”). In order to compare results obtained in the two search modes we defined the Enzyme Unspecificity Factor (EUF) for each protein as the number of unique peptides in the no-enzyme MASCOT search divided by the number of unique peptides in the trypsin MASCOT search. As seen in the table, ZP\_02147451 is the only protein having a higher number of unique peptides (18-fold more) when the “no-enzyme” condition is used. MASCOT search with the “no-enzyme” parameter does not change significantly the results of spectral counting when cellular proteome are analyzed as based on our





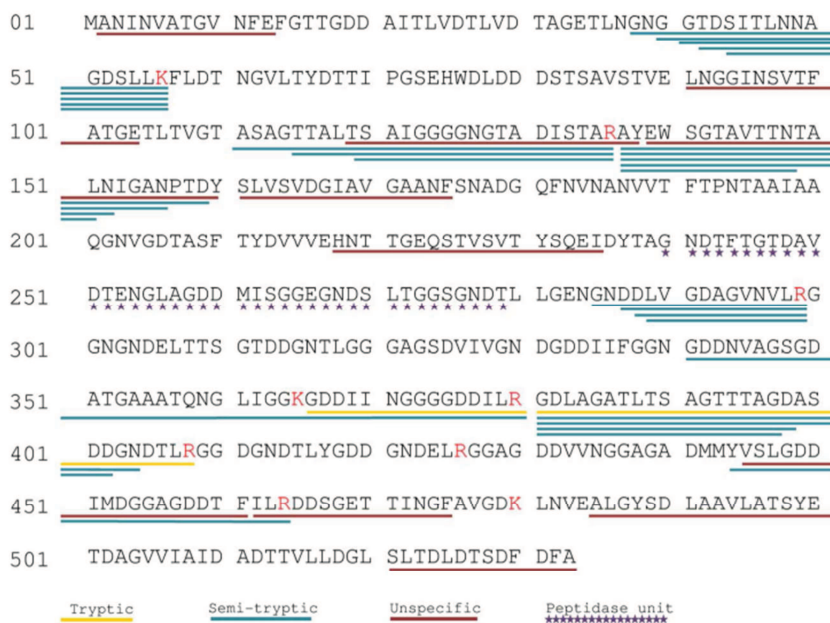
**Figure 2. MS/MS spectrum of the semi-tryptic peptide [288–299] belonging to ZP\_02147451.** The MS/MS spectrum was acquired with a FT/FT procedure with an LTQ-Orbitrap XL hybrid mass spectrometer. The peptide sequence is shown on the top with the collision-induced fragmentation pattern. The *b* and *y* ions are shown in blue and red, respectively. The *y*<sub>11</sub> di-charged ion is labeled in green.  
doi:10.1371/journal.pone.0089691.g002

previous data (Data not shown) from another *Roseobacter*, namely *Roseobacter denitrificans* [35].

### The Recalcitrance of ZP\_02147451 is Caused by an Extreme Lack of Basic Residues in its Sequence

The ZP\_02147451 protein is a RTX-toxin with 533 amino acids that contains only a few basic residues (three lysines and six arginines). These residues are trypsin targets for proteolysis and, therefore, this RTX-like protein will derive in extremely large peptides after digestion. Considering the length of the protein and

the position of these basic residues, only three tryptic peptides could be considered as detectable by the LC-MS/MS procedure with the LTQ-Orbitrap XL hybrid mass spectrometer (*m/z* range from 300 to 1800 th). When reconsidering all the peptides detected of ZP\_02147451, less than half of the cleavages occurred at the C-terminus of arginines (30%) and lysines (12%) in accordance to the trypsin activity. Nevertheless, a large number of cleavages (20%) also arose at the C-terminus of asparagine (N). Other residues like glutamic acid (8%) and aspartic acid (7%) were also frequently found at the C-terminus of the detected peptides (Figure S1). Thus, acidic residues frequently appeared at this



**Figure 3. ZP\_02147451 sequence coverage with non-tryptic, semi-tryptic, and tryptic peptides.** The ZP\_02147451 sequence is represented with its peptidase motif (residues 240 to 279) pointed out with purple stars. Peptides identified with the “no-enzyme” Mascot search are symbolized with a line underlining the sequence. Tryptic, semi-tryptic, and non-tryptic peptides are indicated in yellow, blue and red, respectively.  
doi:10.1371/journal.pone.0089691.g003

**Table 2.** List of proteins identified in the 55 kDa band using the high resolution FT/FT proteomic procedure<sup>1</sup>.

Accession	Functional annotation	Molecular weight (Da)	"Trypsin" as MASCOT search criterion					"No-enzyme" as MASCOT search criterion				
			Protein Mascot Score	Sequence Coverage (%)	Number of unique peptides	Spectral Count	EUFS <sup>2</sup>	Protein Mascot Score	Sequence Coverage (%)	Number of unique peptides	Spectral Count	
ZP_02145205	chaperonin GroEL	57749	1118	50	22	38	68	905	37	15	20	
ZP_02144235	Hemolysin-type calcium-binding protein	115931	685	13	12	48	92	711	15	11	40	
ZP_02146523	oligopeptide/dipeptide ABC transporter*	57342	533	27	10	24	70	440	22	7	12	
ZP_02145689	extracellular solute-binding protein, family 1	64572	354	19	7	12	71	293	13	5	7	
ZP_02146184	extracellular solute-binding protein, family 5	61051	339	17	6	12	83	308	14	5	11	
ZP_02144479	extracellular solute-binding protein, family 5	57617	338	18	6	10	83	302	15	5	6	
ZP_02143510	Ser/Thr protein phosphatase/nucleotidase	56013	263	16	6	10	83	235	14	5	9	
ZP_02143869	phosphate ABC transporter*	37145	221	9	2	3	100	221	9	2	3	
ZP_02144175	aconitate hydratase	96861	232	9	6	9	33	112	3	2	2	
ZP_02147263	outer membrane porin	31831	213	20	5	10	40	115	7	2	5	
ZP_02145456	Flagellar hook-associated protein	50979	196	15	4	4	75	163	10	3	3	
ZP_02145455	S-adenosyl-L-homocysteine hydrolase	45860	195	13	4	4	100	256	17	4	4	
ZP_02143988	Hemolysin-type calcium-binding protein	30261	190	24	3	6	100	188	24	3	5	
ZP_02144038	hypothetical protein RGB5107_15846	51273	152	12	4	8	75	127	7	3	5	
ZP_02147451	Hemolysin-type calcium-binding protein	52624	163	8	2	4	1800	2463	51	36	67	
ZP_02146962	extracellular solute-binding protein, family 1	47415	154	12	3	4	67	125	8	2	2	

<sup>1</sup>Detected with at least three different peptides. The values are from a representative experiment.

\*Periplasmic component.

<sup>2</sup>EUFS stands for Enzyme Unspecificity Factor, defined as the ratio number of unique peptides (no-enzyme MASCOT search) per number of unique peptides (trypsin MASCOT search). doi:10.1371/journal.pone.0089691.t002

position although trypsin could not explain such cleavages. Taken together, the low number of detectable tryptic peptides of the RTX-toxin ZP\_02147451 is the reason why we observed other proteolytic derived peptides. These alternative proteolytic activities are usually rare and masked by the highly abundant tryptic peptides. Furthermore, in standard analysis these peptides would not be considered during the search.

We subjected the whole exoproteome of *Phaeobacter strain DSM 17395* to a novel shotgun analysis with the FT/FT procedure in order to apply a *de novo* sequencing approach and evaluate alternative proteolysis patterns. The data were analyzed with MASCOT using a “no-enzyme” or “trypsin” parameter. **Table 3** shows the comparative list of proteins assigned with at least five different peptides in the “no-enzyme” Mascot analysis mode. Interestingly, with “trypsin” as criterion for the Mascot search, ZP\_02147451 was detected with only one unique peptide (2 spectral counts). The same protein was detected with 12 peptides when the “no-enzyme” parameter was used. ZP\_02147451 was the only protein that exhibited such a large increase of detected peptides when the non-tryptic search mode was used (12-fold, **Table 3**). Noteworthy, we did not find any specific proteolytic pattern in the other proteins because of the small number of extra non-tryptic peptides detected during the analysis.

### RTX-toxins Exhibit Unusually Low Content in Basic Residues

After identifying the RTX-like protein ZP\_02147451 by *de novo* sequencing we further investigated the reason of its recalcitrance to our standard proteomic approach. Within the 533 residue protein we only found nine basic amino acids. Thus, it results refractory to standard trypsin-based proteomic approaches because of the large tryptic peptides generated. These cannot be analyzed properly by mass spectrometers operated in classical conditions. We checked whether this was a specific feature for this protein compared to the whole *Phaeobacter strain DSM 17395* theoretical proteome. **Figure 4** shows the comparison for all the proteins with length above 100 residues. Indeed, ZP\_02147451 is by far unusual in terms of basic residue content, *i.e.* 1 basic residue every 59 residues while the mean value is every 11 residues within the whole theoretical proteome of *Phaeobacter strain DSM 17395*. Noteworthy, we detected that other RTX-toxins encoded on the *Phaeobacter strain DSM 17395* genome are also unusually poor in basic residues. ZP\_02146514, ZP\_02146960, ZP\_02146988, ZP\_02146235, and ZP\_02146693 are large RTX-toxin proteins (1725, 969, 296, 1112, and 937 amino acids respectively) which contain 1 basic residue every 30 residues in average (**Figure 4**). We wondered whether this characteristic is conserved among the other secreted proteins. **Figure 4** shows that the other proteins detected in the exoproteome of *Phaeobacter strain DSM 17395* do not exhibit such a specific trend, although most have a slightly lower RK content compared to the average. This trait is a common issue within the RTX-like proteins coded in *Roseobacter* members. The analysis of the 107 RTX-like proteins coded in 12 other sequenced *Roseobacter* strains revealed one basic residue every 27 amino acids (Data not shown). As an example, *Roseobacter* sp. MED193 exhibits two RTX-like proteins, ZP\_01058551 and ZP\_01057949, with an average tryptic peptide size of 80 and 73 residues, respectively. As a result, the secreted RTX-like proteins are generally refractory to trypsin digestion. We evidenced as illustrated in **Figure 4** the presence in the exoproteome of a large protein with a low ratio KR/polypeptide length through the identification of three distinct peptides. This protein, ZP\_02146938.1, is large (122 kDa) and annotated as the hypothetical protein RGS107\_08210. Its low content in basic residues and large molecular weight may suggest

that it could also be a “RTX-toxin” far-related protein. Indeed, a Blast analysis confirmed such relationship.

### Concluding Remarks

Secreted proteins play an essential role in successful adaptation of bacteria to their environment. These exoproteins are crucial for nutrient import, motility, and biofilm formation. In some cases, they are used by pathogens to adhere to and degrade cell walls from their targets and to suppress defense responses from their hosts [36]. With the development of novel high-throughput proteomic strategies and the use of novel generation of hybrid tandem mass spectrometers, novel insights were recently obtained regarding the secretome from various marine *Roseobacter* strains in several physiological conditions [10,12]. Here, we reported the analysis of the exoproteome of another *Roseobacter* member, namely *Phaeobacter strain DSM 17395*, which is known to produce potent and selective algacides [25]. In this case, we found that the major secreted protein, representing 60% of the total exoproteome, was refractory to classical proteomic approach identification. This challenging issue made us changes our strategy and resort to *de novo* sequencing in order to identify the polypeptide. Interestingly, we found that the abundantly secreted protein was a RTX-like toxin, namely ZP\_02147451, annotated as a hemolysin-type calcium-binding protein. This 533 amino acids polypeptide comprised a domain spanning from amino acids 239 to 279 and shared high similarities to the C-terminal peptidase M10 serralyisin. Noteworthy, we did not find any specific proteolytic pattern for the other proteins because of the small number of extra non-tryptic peptides detected during the analysis. This indicates that no massive proteolysis occurred during exoproteome extraction as previously experienced [7,12,37]. As previously reported, the presence of highly secreted RTX-toxins seems to be a common issue in some *Roseobacter* strains [12]. In this sense, the main component of the exoproteome from *Ruegeria pomeroyi* DSS-3 was the PaxA RTX-toxin, observed whatever the culture condition experimentally tested. While some RTX-like proteins have been well characterized and considered as the main virulence factors in several well-known uropathogenic strains or different pathogenic *Vibrio* species [38] most of these toxin-like proteins remain uncharacterized in nonpathogenic environmental strains. This diverse group of RTX-like secreted proteins encoded in *Roseobacter* members is thought to have a potential effect on other members of the marine community in order to capture organic matter [39]. Interestingly, the secreted RTX-like proteins were depleted from the exoproteome of *R. pomeroyi* DSS-3 only in presence of a natural marine port community indicating a potential effect on its environment [7].

We concluded that the low content in basic residues detected in RTX-like proteins from *Roseobacter* strains could be linked to the function of these proteins rather than to the fact that they are secreted. Because these proteins encompass peptidase domains, their content in basic residues could be constrained to a low level for a higher resistance to the proteolysis of trypsin-like proteases or autoproteolysis as RTX-toxins may autoprocess [40,41]. Because they are secreted to the marine environment in high amounts, at least ZP\_02147451 from *Phaeobacter strain DSM 17395* and PaxA from *R. pomeroyi* DSS-3, their function should be vital for bacteria from the *Roseobacter* clade. The present study shows that usual proteomic parameters for the interpretation of MS/MS data are misleading when exoproteomes are being analyzed. Here, trypsin-specific searches for MS/MS assignment did not identify correctly the ZP\_02147451 RTX-toxin and led to false quantitative results regarding the main exoproteome components. Therefore, we propose for further exoproteome studies to analyze by tandem



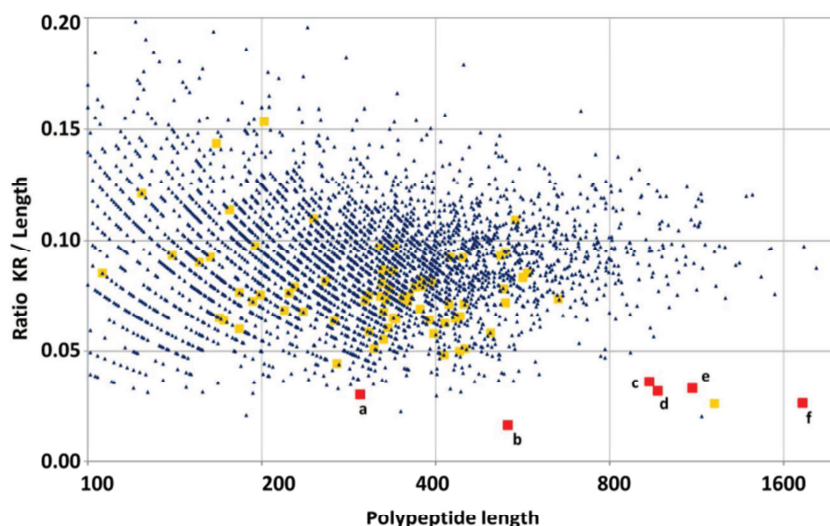
**Table 3.** List of proteins identified in whole exoproteome of *Phaeobacter strain DSM 17395* using the FT/FT procedure<sup>1</sup>.

"Trypsin" as MASCOT search criteria							"No-enzyme" as MASCOT search criteria						
Accession	Functional annotation	Molecular weight (Da)	Protein Mascot Score	Sequence Coverage (%)	Number of unique peptides	Spectral Count	NSAF (%)	EUF <sup>§</sup> (%)	Protein Sequence Mascot Score (%)	Number of unique peptides	Spectral Count	NSAF (%)	
ZP_02144235	Hemolysin-type calcium-binding protein	115932	1109	20	17	70	2.4	88	1043	18	15	59	2.8
ZP_02145485	flagellin-like protein	28065	1057	66	16	63	9.0	106	1157	63	17	47	9.4
ZP_02146962	extracellular solute-binding protein, family 1	47415	888	46	13	34	2.9	92	853	44	12	32	3.8
ZP_02143869	phosphate ABC transporter*	37146	862	49	15	33	3.6	87	816	50	13	23	3.5
ZP_02147451	Hemolysin-type calcium-binding protein	52624	82	3	1	2	0.2	1200	825	21	12	17	1.8
ZP_02145205	chaperonin GroEL	57750	837	45	15	23	1.6	87	839	41	13	19	1.8
ZP_02147263	outer membrane porin	31831	798	34	12	156	19.7	108	848	36	13	100	17.6
ZP_02144302	glutathione synthetase	39420	622	46	10	20	2.0	80	544	34	8	17	2.4
ZP_02146401	phosphonate ABC transporter*	32128	539	49	10	19	2.4	80	513	35	8	14	2.4
ZP_02147202	TRAF dicarboxylate transporter	36347	548	38	9	23	2.5	89	522	33	8	22	3.4
ZP_02143908	TRAF transporter solute receptor	34084	541	35	11	26	3.1	82	498	35	9	16	2.6
ZP_02146891	Amiro acid ABC transporter*	35345	512	41	9	16	1.8	89	471	41	8	14	2.2
ZP_02146522	Hemolysin-type calcium-binding region	31794	509	33	7	21	2.7	100	501	33	7	14	2.5
ZP_02146523	oligopeptide/dipeptide ABC transporter*	57342	499	26	9	18	1.3	78	441	21	7	14	1.4
ZP_02144316	putative iron ABC transporter*	35198	486	42	9	16	1.8	78	428	42	7	14	2.2
ZP_02146543	hypothetical protein RGBS107_00865	47171	479	31	8	17	1.5	75	430	23	6	12	1.4
ZP_02144531	OmpA/MotB	22083	434	82	9	11	2.0	56	325	48	5	6	1.5
ZP_02143879	Extracellular ligand-binding receptor	39919	408	47	8	14	1.4	75	337	36	6	5	1.3
ZP_02146376	putative sugar ABC transporter*	34311	399	41	8	19	2.2	75	345	41	6	12	2.0
ZP_02146184	extracellular solute-binding protein, family 5	61051	377	20	6	10	0.7	100	373	20	6	5	0.8
ZP_02144107	hypothetical protein RGBS107_16191	40145	368	38	7	11	1.1	86	341	28	6	10	1.4
ZP_02146027	EF hand domain protein	17617	363	53	6	10	2.3	67	294	36	4	7	2.2
ZP_02146810	zinc/manganese/iron ABC transporter*	34197	333	15	5	9	1.1	100	328	15	5	7	1.1
ZP_02145961	extracellular solute-binding protein, family 1	48328	322	15	4	6	0.5	125	366	20	5	7	0.8

<sup>1</sup>Detected with at least three different peptides. The values are from a representative experiment.

\*Periplasmic component.

<sup>§</sup>EUF stands for Enzyme Unspecificity Factor, defined as the ratio number of unique peptides (no-enzyme MASCOT search) per number of unique peptides (trypsin MASCOT search).  
doi:10.1371/journal.pone.0089691.t003



**Figure 4. Occurrence of basic residues in proteins from *Phaeobacter* strain DSM 17395.** The graph reports the ratio Lysine (K) and Arginine (R) residues per protein length of all the proteins with length above 100 residues encoded by *Phaeobacter* strain DSM 17395. Proteins are symbolized by a blue triangle. RTX-like proteins and the other exoproteins detected by tandem mass spectrometry are represented by red and yellow squares, respectively. The ZP\_02143988.1 (a), ZP\_02147451.1 (b), ZP\_02144693.1 (c), ZP\_02146960.1 (d), ZP\_02144235.1 (e), and ZP\_02146514.1 (f) RTX-toxins are indicated. Only the ZP\_02143988.1 RTX-like protein has been detected by mass spectrometry. doi:10.1371/journal.pone.0089691.g004

mass spectrometry the proteins cleaved with a combination of different proteases (*e.g.* trypsin, chymotrypsin or EndoGluC proteases) in order to not underestimate, or simply miss, abundantly secreted proteins of interest. As an example, here the experimental proteolysis of the 55 kDa band with chymotrypsin allowed the identification of the ZP\_02147451 RTX-toxin with 701 MS/MS spectra assigned to the protein out of 817 MS/MS spectra recorded, *i.e.* 85% (Table S5 in File S1). Proteolysis with endo glu-C led to the identification of the same protein with 662 assigned MS/MS spectra out of 747 MS/MS spectra recorded, *i.e.* 89 (Table S6 in File S1). In this paper we outlined a strategy that could be systematically used to identify secreted proteins in environmental environments. It would be worth re-analyzing previous exoproteome datasets to check whether the lack of identification of an intense protein was common to several of these studies. However, the lack of the corresponding raw data (SDS-PAGE and MS/MS datasets) prevented such analysis. While RTX-proteins are found numerous in marine bacteria with more than a thousand members, some have been found involved in virulence of uropathogenic *E. coli* revealing strong medical interest [42]. Our results point at the specific structural characteristics of these proteins as basic residues have been counter-selected during evolution.

## Supporting Information

**Figure S1 Relative abundance of C-ter cleaved residues within the whole mass spectrometry-detected ZP\_02147451 peptides.**

## References

- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74.
- Giebel HA, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, et al. (2011) Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J* 5: 8–19.
- Wagner-Dobler I, Biehl H (2006) Environmental biology of the marine Roseobacter lineage. *Annu Rev Microbiol* 60: 255–280.
- Geng H, Belas R (2010) Molecular mechanisms underlying roseobacter-phytoplankton symbioses. *Curr Opin Biotechnol* 21: 332–338.
- Armengaud J (2013) Microbiology and proteomics, getting the best of both worlds! *Environ Microbiol* 15: 12–23.
- Armengaud J, Hartmann EM, Bland C (2013) Proteogenomics for environmental microbiology. *Proteomics* 13: 2731–42.

(TIF)

**File S1 Table S1, List of tryptic peptides identified in the whole exoproteome from *Phaeobacter* strain DSM 17395 using a standard shotgun procedure. Table S2, List of proteins identified in the whole exoproteome from *Phaeobacter* strain DSM 17395 using a standard shotgun procedure. Table S3, List of tryptic peptides identified in the 55 kDa SDS-PAGE band of the exoproteome from *Phaeobacter* strain DSM 17395 using a standard shotgun procedure. Table S4, List of peptides identified in the 55 kDa band of the exoproteome from *Phaeobacter* strain DSM 17395 using the FT/FT procedure and with "no-enzyme" as MASCOT search criteria. Table S5, List of peptides identified in the 55-kDa band of the exoproteome from *Phaeobacter* strain DSM 17395 after chymotrypsin digestion using the FT/FT procedure. Table S6, List of peptides identified in the 55-kDa band of the exoproteome from *Phaeobacter* strain DSM 17395 after endo-GluC digestion using the FT/FT procedure.**

(XLSX)

## Author Contributions

Conceived and designed the experiments: ED JACO JA. Performed the experiments: ED JACO JA. Analyzed the data: ED JACO JA. Contributed reagents/materials/analysis tools: ED JACO JA. Wrote the paper: ED JACO JA.

7. Christie-Oleza JA, Fernandez B, Nogales B, Bosch R, Armengaud J (2012) Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. *ISME J* 6: 124–135.
8. Clair G, Armengaud J, Dupont C (2012) Restricting fermentative potential by proteome remodeling: an adaptive strategy evidenced in *Bacillus cereus*. *Mol Cell Proteomics* 11: M111 013102.
9. Armengaud J, Christie-Oleza JA, Clair G, Malard V, Dupont C (2012) Exoproteomics: exploring the world around biological systems. *Expert Rev Proteomics* 9: 561–575.
10. Christie-Oleza JA, Armengaud J (2010) In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. *Mar Drugs* 8: 2223–2239.
11. Clair G, Roussi S, Armengaud J, Dupont C (2010) Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. *Mol Cell Proteomics* 9: 1486–1498.
12. Christie-Oleza JA, Pina-Villalongo JM, Bosch R, Nogales B, Armengaud J (2012) Comparative proteogenomics of twelve roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol Cell Proteomics* 11: M111 013110.
13. Chagnot C, Zorgani MA, Astruc T, Desvaux M (2013) Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol* 4: 303.
14. Wohlbrand L, Trautwein K, Rabus R (2013) Proteomic tools for environmental microbiology—a roadmap from sample preparation to protein identification and quantification. *Proteomics* 13: 2700–2730.
15. Martens T, Heidorn T, Pukall R, Simon M, Tindall BJ, et al. (2006) Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1999 as *Marinovum algicola* gen. nov., comb. nov., and emended descriptions of the genera *Roseobacter*, *Ruegeria* and *Leisingera*. *Int J Syst Evol Microbiol* 56: 1293–1304.
16. Ruiz-Ponte C, Cilia V, Lambert C, Nicolas JL (1998) *Roseobacter gallaeciensis* sp. nov., a new marine bacterium isolated from rearings and collectors of the scallop *Pecten maximus*. *Int J Syst Bacteriol* 48 Pt 2: 537–542.
17. Yoon JH, Kang SJ, Lee SY, Oh TK (2007) *Phaeobacter daeponensis* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. *Int J Syst Evol Microbiol* 57: 856–861.
18. Zhang DC, Li HR, Xin YH, Liu HC, Chi ZM, et al. (2008) *Phaeobacter arcticus* sp. nov., a psychrophilic bacterium isolated from the Arctic. *Int J Syst Evol Microbiol* 58: 1384–1387.
19. Druppel K, Hensler M, Trautwein K, Kossmehl S, Wohlbrand L, et al. (2014) Pathways and substrate-specific regulation of amino acid degradation in *Phaeobacter inhibens* DSM 17395 (archetype of the marine *Roseobacter* clade). *Environ Microbiol* 16: 218–38.
20. Kossmehl S, Wohlbrand L, Druppel K, Feenders C, Blasius B, et al. (2013) Subcellular protein localization (cell envelope) in *Phaeobacter inhibens* DSM 17395. *Proteomics* 13: 2743–2760.
21. Zech H, Hensler M, Kossmehl S, Druppel K, Wohlbrand L, et al. (2013) Dynamics of amino acid utilization in *Phaeobacter inhibens* DSM 17395. *Proteomics* 13: 2869–2885.
22. Zech H, Hensler M, Kossmehl S, Druppel K, Wohlbrand L, et al. (2013) Adaptation of *Phaeobacter inhibens* DSM 17395 to growth with complex nutrients. *Proteomics* 13: 2851–2868.
23. Berger M, Neumann A, Schulz S, Simon M, Brinkhoff T (2011) Tropodithietic acid production in *Phaeobacter gallaeciensis* is regulated by N-acetyl homoserine lactone-mediated quorum sensing. *J Bacteriol* 193: 6576–6585.
24. Brinkhoff T, Giebel HA, Simon M (2008) Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. *Arch Microbiol* 189: 531–539.
25. Seyedsayamdost MR, Carr G, Kolter R, Clardy J (2011) Roseobactin: small molecule modulators of an algal-bacterial symbiosis. *J Am Chem Soc* 133: 18343–18349.
26. Seyedsayamdost MR, Case RJ, Kolter R, Clardy J (2011) The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat Chem* 3: 331–335.
27. Buddhuhs N, Pradella S, Goker M, Pauker O, Pukall R, et al. (2013) Molecular and phenotypic analyses reveal the non-identity of the *Phaeobacter gallaeciensis* type strain deposits CIP 105210T and DSM 17395. *Int J Syst Evol Microbiol* 63: 4340–4349.
28. Dedieu A, Gaillard JC, Pourcher T, Darrouzet E, Armengaud J (2011) Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy. *J Biol Chem* 286: 259–269.
29. de Groot A, Dulermo R, Ortet P, Blanchard L, Guerin P, et al. (2009) Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet* 5: e1000434.
30. Dupierri V, Masselon C, Court M, Kieffer-Jaquinod S, Bruley C (2009) A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* 25: 1980–1981.
31. Liu H, Sadygov RG, Yates JR 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76: 4193–4201.
32. Zivanovic Y, Armengaud J, Lagorce A, Leplat C, Guerin P, et al. (2009) Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the *Archaea*. *Genome Biol* 10: R70.
33. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, et al. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A* 103: 18928–18933.
34. Baudet M, Ortet P, Gaillard JC, Fernandez B, Guerin P, et al. (2010) Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons. *Mol Cell Proteomics* 9: 415–426.
35. Bland C, Bellanger L, Armengaud J (2014) Magnetic Immunoaffinity Enrichment for Selective Capture and MS/MS Analysis of N-Terminal-TMPP-Labeled Peptides. *J Proteome Res*. in press.
36. Preston GM, Studholme DJ, Caldelari I (2005) Profiling the secretomes of plant pathogenic Proteobacteria. *FEMS Microbiol Rev* 29: 331–360.
37. Clair G, Lorphelin A, Armengaud J, Dupont C (2013) OhrRA functions as a redox-responsive system controlling toxinogenesis in *Bacillus cereus*. *J Proteomics* 94: 527–539.
38. Linhartova I, Bumba L, Masin J, Basler M, Osicka R, et al. (2010) RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* 34: 1076–1112.
39. Moran MA, Belas R, Schell MA, Gonzalez JM, Sun F, et al. (2007) Ecological genomics of marine *Roseobacters*. *Appl Environ Microbiol* 73: 4559–4569.
40. Satchell KJ (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu Rev Microbiol* 65: 71–90.
41. Sheahan KL, Cordero CL, Satchell KJ (2007) Autoprocessing of the *Vibrio cholerae* RTX toxin by the cysteine protease domain. *EMBO J* 26: 2552–2561.
42. Vigil PD, Alteri CJ, Mobley HL (2011) Identification of in vivo-induced antigens including an RTX family exoprotein required for uropathogenic *Escherichia coli* virulence. *Infect Immun* 79: 2335–2344.



# Proteogenomic Biomarkers for Identification of *Francisella* Species and Subspecies by Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometry

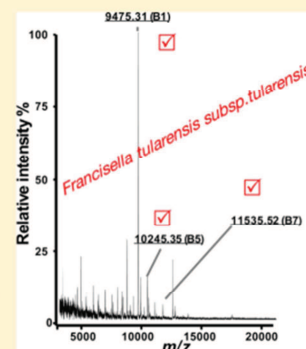
Emie Durighello,<sup>†</sup> Laurent Bellanger,<sup>‡</sup> Eric Ezan,<sup>†,‡</sup> and Jean Armengaud<sup>\*,†</sup>

<sup>†</sup>CEA Marcoule, DSV, IBEB, Laboratoire de Biochimie des Systèmes Perturbés, Bagnols-sur-Cèze, F-30207, France

<sup>‡</sup>CEA Marcoule, DSV, IBEB, Laboratoire d'Ingénierie Cellulaire et Biotechnologie, Bagnols-sur-Cèze, F-30207, France

## Supporting Information

**ABSTRACT:** *Francisella tularensis* is the causative agent of tularemia. Because some *Francisella* strains are very virulent, this species is considered by the Centers for Disease Control and Prevention to be a potential category A bioweapon. A mass spectrometry method to quickly and robustly distinguish between virulent and nonvirulent *Francisella* strains is desirable. A combination of shotgun proteomics and whole-cell matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry on the *Francisella tularensis* subsp. *holarctica* LVS defined three protein biomarkers that allow such discrimination: the histone-like protein HU form B, the 10 kDa chaperonin Cpn10, and the 50S ribosomal protein L24. We established that their combined detection by whole-cell MALDI-TOF spectrum could enable (i) the identification of *Francisella* species, and (ii) the prediction of their virulence level, i.e., gain of a taxonomical level with the identification of *Francisella tularensis* subspecies. The detection of these biomarkers by MALDI-TOF mass spectrometry is straightforward because of their abundance and the absence of other abundant protein species closely related in terms of  $m/z$ . The predicted molecular weights for the three biomarkers and their presence as intense peaks were confirmed with MALDI-TOF/MS spectra acquired on *Francisella philomiragia* ATCC 25015 and on *Francisella tularensis* subsp. *tularensis* CCUG 2112, the most virulent *Francisella* subspecies.



Tularemia is an infectious bacterial disease caused by some subspecies of *Francisella tularensis*. Because of its virulence characteristics, its ease of dissemination, its generally low normal prevalence, and reported cases of asymptomatic infection and fatal disease, *Francisella tularensis* is categorized by the U.S. Centers for Disease Control and Prevention as a category A bioterrorism agent.<sup>1,2</sup> The emergence of new *Francisella* pathogens, such as those known to be the cause of numerous diseases in aquaculture,<sup>3</sup> could be of great concern if their mechanisms of immune subversion are modified by synthetic biology. Great concerns for rapid, simple to perform, robust, low-cost, and high-throughput detection have been raised in recent years in parallel with global political instability.<sup>4–7</sup> Improving detection through the detection of specific, robust biomarkers is of utmost importance.

The *Francisella* genus comprises three different characterized species: *Francisella tularensis*, *Francisella novicida*, and *Francisella philomiragia*.<sup>8</sup> The first species is further subdivided into three subspecies on the basis of phenotypic traits. While *F. novicida* and *F. philomiragia* species have always been associated with a low level of virulence, *F. tularensis* subsp. *tularensis* (type A) exhibits a very high level of virulence, translating into a human mortality rate reaching 30% when antibiotic treatment is not available.<sup>9</sup> *F. tularensis* subsp. *holarctica* (type B) and *F. tularensis* subsp. *mediasiatica* have high virulence but almost no cases of human mortality have been reported.<sup>8</sup> Genotyping of *F. tularensis* has shown its low sequence diversity, with

pairwise comparison of average nucleotide identities above 99.2% across subspecies.<sup>10</sup> Whole-cell matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry is now a routine technique in many microbiology laboratories and was developed for *Francisella*.<sup>11</sup> The mass fingerprint of the most abundant, basic, and low molecular weight proteins is compared with the profiles previously recorded for similar bacteria.<sup>12–14</sup> However, only a restricted number of bacteria, mostly known pathogens, have been recorded to date and their  $m/z$  profiles deposited in databases. Because of the huge reservoir of diversity on Earth, and the emergence of new pathogens, the whole-cell MALDI-TOF method should be further complemented with novel approaches based on specific biomarkers.<sup>15–17</sup> Specific  $m/z$  values in the mass spectrometry (MS) spectra are required to further distinguish closely related subspecies, as clustering methods do not perform well if spectra have a low number of signals or are recorded for a sample consisting of a mixture of bacteria.<sup>18</sup> Interestingly, immunoaffinity and MALDI-TOF detection have been combined for the sensitive detection of *F. tularensis*,<sup>19</sup> but this method requires a proteolysis step.

Received: April 30, 2014

Accepted: September 12, 2014

Published: September 12, 2014



Here, we identified the proteins contributing to the whole-cell MALDI-TOF profile of *F. tularensis* bacteria and predicted the  $m/z$  values for homologues found in the whole set of *Francisella* characterized to date in terms of genome sequencing. Proteogenomics, the combination of genomic and proteomic data to better annotate key proteins and their corresponding genes,<sup>20,21</sup> is helpful to revisit the definition of biomarkers for environmental microbiology.<sup>15,22</sup> We selected a trio of specific biomarkers that could be used to identify *Francisella* subspecies and to quickly predict the level of pathogenicity of the sample. For this, we analyzed *Francisella tularensis* subsp. *holarctica* LVS by whole-cell MALDI-TOF. Concomitantly, we identified, with a shotgun approach, the main components of its low molecular weight proteome after trypsin proteolysis and nano liquid chromatography–tandem mass spectrometry (LC–MS/MS). Among the 12 proteins identified as abundant in the MALDI-TOF profile, three were found to be relatively conserved among the *Francisella* genus, but their slight differences in terms of sequences allowed the distinction of subspecies. We verified the accuracy of our predictions by recording and analyzing whole-cell MALDI-TOF spectra for two other *Francisella* species and subspecies, one being the most virulent known subspecies of *Francisella*.

## EXPERIMENTAL SECTION

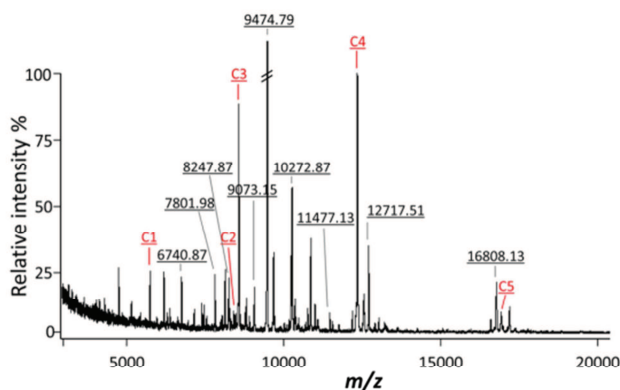
**Francisella Culture and Inactivation.** *F. tularensis* subsp. *holarctica* LVS and *F. philomiragia* ATCC 25015 were manipulated in BSL2 and then inactivated. *F. tularensis* subsp. *tularensis* CCUG 2112 was manipulated in BSL3 and then inactivated. All the experiments regarding these strains were performed according to the legal rules for their handling and under legal authorization delivered by the French regulatory agency, the National Security Agency for Medicines and Health Products (ANSM). *F. tularensis* subsp. *holarctica* LVS, *F. philomiragia* ATCC 25015, and *F. tularensis* subsp. *tularensis* CCUG 2112 were grown in 40 mL of brain heart infusion (BHI) liquid broth and 1% PolyViteX (bioMérieux) at 37 °C. Cells were harvested by centrifugation at 2 500g for 10 min at 20 °C when the stationary phase was reached. Inactivation was performed with 100  $\mu$ L of 80% trifluoroacetic acid (Sigma-Aldrich) per 10 mg of bacteria. Inactivated samples were incubated for 30 min at 4 °C under 180 rpm agitation, supplemented with 150  $\mu$ L of 67% acetonitrile and then centrifuged at 13 000g for 5 min. Supernatants containing the soluble proteins were systematically checked for inactivation by inoculating a chocolate agar plate (bioMérieux) with 100  $\mu$ L of inactivated solution and incubating this plate for 6 days at 37 °C. Inactivated protein samples were removed from the BSL2 or BSL3 facilities for MALDI-TOF acquisition only after validation that no colonies had grown after the 6-day period.

**Mass Spectrometry.** After TFA inactivation, protein samples were desalted using ZipTip<sub>C18</sub> (Millipore) and spotted onto an MTP 384 polished steel target plate (Bruker) in the presence of protein calibration standard I (Bruker) and sinapinic acid as matrix, as previously recommended.<sup>15,16</sup> The Biflex IV MALDI-TOF instrument (Bruker) was operated in the 3 000–20 000  $m/z$  range. Four spectra were selected after internal calibration, with at least four of the six standards detected with an error tolerance below 50 ppm. Protein samples were resolved by SDS-PAGE, and the low molecular weight proteome (<20 kDa) was treated as previously described.<sup>23,24</sup> NanoLC–MS/MS was performed with a high-resolution LTQ-Orbitrap XL hybrid mass spectrometer

(ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings), as previously described.<sup>25</sup> MS/MS spectra were assigned with standard criteria with the MASCOT 2.3.02 program (Matrix Science) using the complete list of annotated coding sequences from *Francisella tularensis* subsp. *holarctica* LVS (NCBI downloaded, January 10, 2012). Peptides were filtered with a  $p$ -value below 0.05.

## RESULTS AND DISCUSSION

**Whole-Cell MALDI-TOF and NanoLC–MS/MS of *Francisella tularensis* subsp. *holarctica* LVS.** We acquired the MALDI-TOF profile of *Francisella tularensis* subsp. *holarctica* LVS in the 3 000 to 20 000  $m/z$  range. For this, proteins were mixed with a standard including six known mass values to enhance mass accuracy through internal calibration. Figure 1 shows a representative spectrum from which 100 of

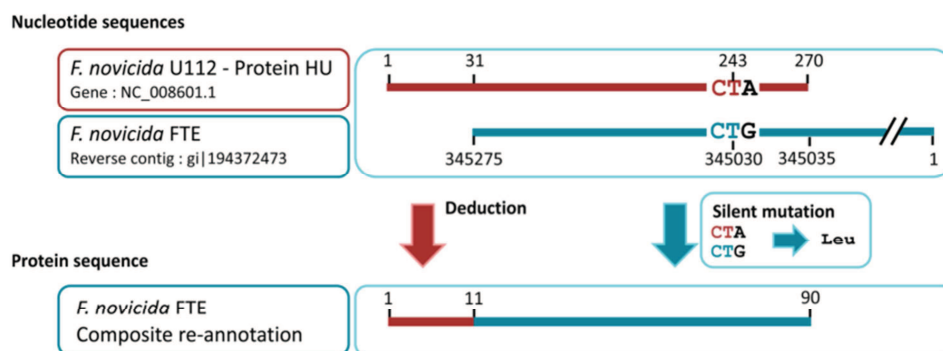


**Figure 1.** Whole cell MALDI-TOF spectrum obtained with *F. tularensis* subsp. *holarctica* LVS cells. The spectrum was internally calibrated using the Protein Calibration Standard I from Bruker Daltonics. Calibrants are indicated by C1, C2, C3, C4, and C5 labels. The  $m/z$  values of the most abundant protonated proteins are indicated.

the most intense peaks could be extracted with the centroid peak detection algorithm. After internal calibration, the standard deviation was about 25 ppm. We assayed the stability of MALDI-TOF profiles for *Francisella tularensis* subsp. *holarctica* LVS cells grown on chocolate agar + PolyViteX (PVX) plates after 1, 2, 3, 6, 7, 8, 9, and 10 days incubation at 37 °C (Figure S1, Supporting Information). We did not observe significant differences, except lower signal intensity after 1 week of incubation. We analyzed and quantified by shotgun proteomics, the 3–20 kDa proteome. A total of 535 spectra could be assigned to 282 specific peptides (Table S1, Supporting Information), which validated the presence of 71 proteins (Table S2, Supporting Information). In terms of protein abundance estimated by spectral counting, ribosomal proteins accounted for 50% of the detected proteome (228 spectra out of 454). As in various bacteria,<sup>15,26</sup> the DNA-binding protein HU was abundantly detected (36 spectra, 8%).

**Identification of 12 Major Proteins Detected by MALDI-TOF.** In order to identify specific biomarkers for *F. tularensis* subsp. *holarctica* LVS, the 100 most intense  $m/z$  peaks identified from four MALDI-TOF spectra were selected. A total of 44  $m/z$  values were present in at least three out of four independently acquired spectra. These  $m/z$  values were converted into theoretical mass values, assuming that they could correspond to monoprotonated or diprotonated ions





**Figure 2.** Proteogenomic annotation of the Histone-like protein HU from *F. novicida* FTE strain. A partial coding sequence for the HU protein has been located at one of the extremities of a genomic contig of *F. novicida* FTE genome by means of a Blastn search with HU nucleotide sequence from a closely related strain, namely, *F. novicida* U112d. A unique silent mutation at the nucleotide level has been detected resulting in a protein sequence of 80 amino acids similar to its closest full-length homologues (90 amino acids). We consider that it is reasonable to assume that the HU protein exists in *F. novicida* FTE with exactly the same protein sequence as *F. novicida* U112 homologue, resulting in the same molecular weight. The composite resulting sequence is indicated.

(Table S3, Supporting Information). We compared this list of average masses with those of proteins abundantly detected by the proteomic nanoLC–MS/MS approach, taking into consideration all possible maturations of the polypeptides, such as acetylation (+ 42.0 Da) and/or methionine loss (– 131.2 Da) at the N-terminus of each polypeptide. A total of 18 *m/z* values were assigned to proteins with a tolerance below 264 ppm. This set was restricted to 15 *m/z* values (12 proteins) when the tolerance was set below 105 ppm (Table S4, Supporting Information). Seven of these proteins were ribosomal proteins. The most intense *m/z* signal on the four MALDI-TOF spectra was assigned to the histone-like protein HU form B (average theoretical *m/z* value, 9474.9). Four proteins have their initial methionine processed and two are N-terminally acetylated. The deciphering of these post-translational modifications is important for predicting the theoretical molecular weight of these biomarkers.

**Proteogenomic Prediction of the 12 Identified Proteins in All Sequenced *Francisella*.** A total of 24 different *Francisella* strains have been sequenced (April 23, 2013) and their genomes released: six *Francisella tularensis* subsp. *tularensis*, seven *Francisella tularensis* subsp. *holarctica*, one *Francisella tularensis* subsp. *mediasiatica*, seven *Francisella novicida*, two *Francisella philomiragia*, and *Francisella* sp. TX077308. We identified the homologues of the 12 possible biomarkers (Table S4, Supporting Information). Because some genomes are only partially sequenced or incorrectly annotated, some homologous proteins could not be identified, although they are conserved among numerous, if not all, bacteria. We searched the missing genes and corrected some annotation errors by means of nucleotide blast, taking as reference the corresponding *Francisella tularensis* subsp. *holarctica* LVS gene sequences. A total of nine genes could be manually corrected. Because some of the new genes are located at the end of a nucleotide contig obtained after partial assembly of sequencing reads, it is impossible to directly establish the molecular weight of the corresponding protein. When no, or silent, nucleotide mutations were found within the known partial sequence, we predicted their probable exact sequences based on the conservation of their sequence among the subspecies and their maturation, if any, as established for the reference proteome. Figure 2 shows the example of the HU protein identified encoded in the *F. novicida* FTE genome and

previously unannotated (NZ\_ABSS01000000). The HU protein is essential in bacteria, but the corresponding gene remained to be discovered in this strain, being expected to be conserved in terms of sequence. In this case, we found that the nucleotide sequence corresponding to amino acids 31 to 273 (90% of the full-length sequence) is at a contig extremity. While this sequence corresponds exactly to the homologous sequence of *F. novicida* U112 strain (NC\_008601.1), except for a silent mutation, it is reasonable to assume that the HU protein exists in *F. novicida* FTE with exactly the same protein sequence as the *F. novicida* U112 homologue, resulting in the same molecular weight by composite proteomics. In some cases, as for the 30S ribosomal protein S16 (protein B4), bacteria from different subspecies have homologues with the same mass values while the amino acid sequences are different (Table S4, Supporting Information). Indeed, in positions 61 and 82 of the B4 protein sequences, residues A and V are found for *philomiragia* strains, while they are V and A, respectively, for the *novicida* strain GA99-3548. This sequence modification does not lead to a mass value shift, so the two proteins cannot be distinguished by MALDI-TOF mass spectrometry. Finally, on the basis of the molecular weights predicted for the 12 proteins from the *Francisella novicida* 3523 strain, the taxonomy of this strain is challenged. A total of eight mass value differences, as compared with the 12 proteins from the whole set of strains, were noted, while the differences among closely related strains do not usually exceed two molecular weights. However, the 16S RNA (NC\_017449.1; gi|387823583:415215-416732) clearly indicates the proximity of *Francisella novicida* 3523 to *Francisella* strains.

**Defining Biomarkers within the *Francisella* Genus to Distinguish Pathogen and Nonpathogen Groups.** We determined which of these proteins could be specific to the *Francisella* genus, the *tularensis* species, or indicators for subspecies or strain levels when considering their global masses. Some protein sequences are well conserved among the different *Francisella* species, such as the 50S ribosomal protein L30 (Table S4, Supporting Information). For this protein, only two different sequences were listed among the 24 different *Francisella* strains sequenced. Other protein sequences are significantly different, such as the B9 and B12 proteins annotated as hypothetical proteins, with seven and six different sequences, respectively. These proteins, although detected in



high abundance in the whole-cell MALDI-TOF spectra of *Francisella tularensis* subsp. *holarctica* LVS strain, cannot be considered as valuable biomarkers because of the diversity of their sequences. A single protein mass cannot be specific for a subspecies of *Francisella tularensis*. However, a combination of three proteins allows the identification of the species and subspecies of a strain belonging to the *Francisella* genus. These proteins are the histone-like protein HU form B (B1), the 10 kDa chaperonin Cpn10 (B5), and the 50S ribosomal protein L24 (B7). Table 1 presents the minimal set of masses for these

**Table 1. Predicted MALDI-TOF  $m/z$  Values for the Three Monocharged Biomarkers According to Their Species and Subspecies**

species	subspecies	B1	B5	B7
<i>F. tularensis</i>	<i>holarctica</i>	9474.9	10272.9	11477.3
	<i>tularensis</i>	9474.9	10245.5	11535.4
	<i>mediasiatica</i>	9474.9	10245.5	11477.3
<i>F. novicida</i>		9444.9	10245.5	11477.3
		9417.8	10176.7	11493.3
<i>F. philomiragia</i>		9392.8	10248.7	11449.3
<i>F. TX077308</i>		9392.8	10249.7	11435.3

three proteins, allowing discrimination between *Francisella* species and *Francisella tularensis* subspecies, based on our predictions. As an example, the identification of  $m/z$  masses corresponding to proteins with a global average mass of 9 473.9, 10 271.9, and 11 476.3 Da should correspond to *F. tularensis* subsp. *holarctica* cells. Moreover, the B1, B5, and B7 values should be exclusive when the analyzed cells correspond to a unique strain.

**Validation of the Predicted Biomarkers with *F. philomiragia* ATCC 25015 and *F. tularensis* subsp. *tularensis* CCUG 2112 Strains.** We checked whether our predictions were correct with two strains. The experimental  $m/z$  values of the three predicted biomarkers, B1, B5, and B7, in the monoprotonated form for *Francisella philomiragia* ATCC 25015 are 9 392.8, 10 248.7, and 11 449.3 (Figure S2, Supporting Information). These assignments agree well with our predictions within a 63 ppm tolerance range. Moreover, the HU protein biomarker (B1) was the most intense  $m/z$  peak while the two other biomarkers also contributed to intense  $m/z$  signals. We also recorded the whole-cell MALDI-TOF spectrum of *Francisella tularensis* subsp. *tularensis* CCUG 2112 (Figure S3, Supporting Information). The peaks at  $m/z$  9 475.3, 10 245.4, and 12 535.5 were within our predictions in a 42 ppm tolerance range. As expected, no  $m/z$  values predicted for the other *Francisella* species could be detected for the three biomarkers listed in Table 1. On the basis of these results, the use of these three specific biomarkers appears to be robust for the identification of *Francisella* species and *F. tularensis* subspecies.

**Relevance of the Three Biomarkers Regarding Global Database Searches.** Because of the huge diversity of bacteria on Earth, many different bacteria could exhibit the three  $m/z$  values, while not being *Francisella* strains.<sup>27</sup> For example, we performed the search with the TagIdent tool for the three molecular weight values (9 473.9, 10 272.9, and 11 477.3) of *F. tularensis* subsp. *holarctica* biomarkers, taking into account a 100 ppm mass accuracy and an isoelectric point (pI) in the 8.5–14 range, and looking among the entire known repertoire of bacterial protein sequences. We observed that 1595, 1514,

and 913 proteins, respectively, could match to these three values in the Swiss-Prot and TrEMBL databases (release 2014\_07 of July 9, 2014). Using the combination of the three masses, 54 bacteria can be possibly identified. However, this theoretical work does not take into consideration that only highly produced proteins, such as the three biomarkers, will be detected by whole-cell MALDI-TOF. As a consequence, a much lower number of bacteria could potentially exhibit three  $m/z$  intense signals matching the three biomarkers. Such an approach is more discriminative than peptide-based biomarkers.<sup>28</sup>

## CONCLUSIONS

We proposed a set of three  $m/z$  values to be checked for a quick classification of a *Francisella* strain by whole-cell MALDI-TOF mass spectrometry. The histone-like protein HU form B, the 10 kDa chaperonin Cpn10, and the 50S ribosomal protein L24 have  $m/z$  values that allow distinction between the different *Francisella* species currently known as well as the subspecies from *Francisella tularensis*. Because ribosomal and house-keeping proteins in general are often highly abundant under normal culturing conditions, the three biomarkers are expected to be detectable whatever the growth conditions. Because the set of *Francisella* genomes covers this genus well and the selected proteins are among the most conserved, our prediction should be robust if new *Francisella* genomes are sequenced. These can be easily re-evaluated by proteogenomics when further taxonomic or genomic updates are available. We recommend including other genera-specific biomarkers and carrying out additional analysis to confirm the presence of a *Francisella* strain, such as isolation on different nutritive broths and further molecular characterization.

## ASSOCIATED CONTENT

### Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [jean.armengaud@cea.fr](mailto:jean.armengaud@cea.fr). Phone: +33 (0)4 66 79 68 02. Fax: +33 (0)4 66 79 19 05.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Commissariat à l'Energie Atomique et aux Energies Alternatives and the Agence Nationale de la Recherche (Grant ANR-11-SECU-0007). We would like to express our sincere gratitude to Edward R. Moore (University of Göteborg, Sweden) for strain assistance as well as Jean-Charles Gaillard, Philippe Guérin, and Marie-Anne Roncato (CEA) for technical assistance.

## REFERENCES

- (1) Dennis, D. T.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Friedlander, A. M.; Hauer, J.; Layton, M.; Lillibridge, S. R.; McDade, J. E.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Tonat, K.; Working Group on Civilian Biodefense. *JAMA, J. Am. Med. Assoc.* **2001**, *285*, 2763–2773.
- (2) Rotz, L. D.; Khan, A. S.; Lillibridge, S. R.; Ostroff, S. M.; Hughes, J. M. *Emerg. Infect. Dis.* **2002**, *8*, 225–230.

- (3) Birkbeck, T. H.; Feist, S. W.; Verner-Jeffreys, D. W. *J. Fish Dis.* **2011**, *34*, 173–187.
- (4) Demirev, P. A.; Fenselau, C. *Annu. Rev. Anal. Chem.* **2008**, *1*, 71–93.
- (5) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157–171.
- (6) Ho, Y.-P.; Reddy, P. M. *Clin. Chem.* **2010**, *56*, 525–536.
- (7) Sandrin, T. R.; Goldstein, J. E.; Schumaker, S. *Mass Spectrom. Rev.* **2013**, *32*, 188–217.
- (8) Keim, P.; Johansson, A.; Wagner, D. M. Molecular ecology, epidemiology and evolution of *Francisella*. In *Francisella Tularensis: Biology, Pathogenicity, Epidemiology, and Biodefense*; Annals of the New York Academy of Sciences, Vol. 1105; Kwaik, Y. A., Ed.; Blackwell Pub.: Boston, MA, 2007; pp 30–66.
- (9) Bulut, O. C.; Dyckhoff, G.; Splettstoesser, W.; Nemeth, J.; Klauschen, F.; Penzel, R.; Plinkert, P. K.; Simon, C.; Weichert, W.; Stenzinger, A. *Int. J. Surg. Pathol.* **2013**, *21*, 76–81.
- (10) Johansson, A.; Petersen, J. M. *J. AOAC Int.* **2010**, *93*, 1930–1943.
- (11) Seibold, E.; Maier, T.; Kostrzewa, M.; Zeman, E.; Splettstoesser, W. *J. Clin. Microbiol.* **2010**, *48*, 1061–1069.
- (12) Sauer, S.; Kliem, M. *Nat. Rev. Microbiol.* **2010**, *8*, 74–82.
- (13) Kliem, M.; Sauer, S. *Curr. Opin. Microbiol.* **2012**, *15*, 397–402.
- (14) Fenselau, C. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1161–1166.
- (15) Christie-Oleza, J. A.; Maria Pina-Villalongo, J.; Guerin, P.; Miotello, G.; Bosch, R.; Nogales, B.; Armengaud, J. *Environ. Microbiol.* **2013**, *15*, 133–147.
- (16) Hartmann, E. M.; Durighello, E.; Pible, O.; Nogales, B.; Beltrametti, F.; Bosch, R.; Christie-Oleza, J. A.; Armengaud, J. *Mar. Genomics* **2014**, DOI: 10.1016/j.margen.2014.04.003.
- (17) Holland, R. D.; Duffy, C. R.; Rafii, F.; Sutherland, J. B.; Heinze, T. M.; Holder, C. L.; Voorhees, K. J.; Lay, J. O., Jr. *Anal. Chem.* **1999**, *71*, 3226–3230.
- (18) Sauer, S.; Freiwald, A.; Maier, T.; Kube, M.; Reinhardt, R.; Kostrzewa, M.; Geider, K. *PLoS One* **2008**, *3*, e2843.
- (19) Jiang, J.; Parker, C. E.; Fuller, J. R.; Kawula, T. H.; Borchers, C. H. *Anal. Chim. Acta* **2007**, *605*, 70–79.
- (20) Armengaud, J. *Curr. Opin. Microbiol.* **2009**, *12*, 292–300.
- (21) Armengaud, J.; Hartmann, E. M.; Bland, C. *Proteomics* **2013**, *13*, 2731–2742.
- (22) Armengaud, J. *Environ. Microbiol.* **2013**, *15*, 12–23.
- (23) Clair, G.; Roussi, S.; Armengaud, J.; Duport, C. *Mol. Cell. Proteomics* **2010**, *9*, 1489–1498.
- (24) Clair, G.; Armengaud, J.; Duport, C. *Mol. Cell. Proteomics* **2012**, *11*, M111.013102.
- (25) de Groot, A.; Dulerio, R.; Ortet, P.; Blanchard, L.; Guerin, P.; Fernandez, B.; Vacherie, B.; Dossat, C.; Jolivet, E.; Siguier, P.; Chandler, M.; Barakat, M.; Dedieu, A.; Barbe, V.; Heulin, T.; Sommer, S.; Achouak, W.; Armengaud, J. *PLoS Genetics* **2009**, *5*, e1000434.
- (26) Toueille, M.; Mirabella, B.; Guerin, P.; de la Tour, C. B.; Boisnard, S.; Hong Ha, N.; Blanchard, L.; Servant, P.; de Groot, A.; Sommer, S.; Armengaud, J. *J. Proteomics* **2012**, *75*, 2588–2600.
- (27) Fenselau, C.; Russell, S.; Swatkoski, S.; Edwards, N. *Eur. J. Mass Spectrom.* **2007**, *13*, 35–39.
- (28) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasatolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. *Anal. Chem.* **2000**, *72*, 1112–1118.



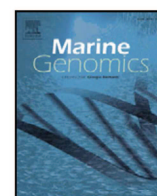






Contents lists available at ScienceDirect

## Marine Genomics

journal homepage: [www.elsevier.com/locate/margen](http://www.elsevier.com/locate/margen)

## Proteomics meets blue biotechnology: A wealth of novelties and opportunities

Erica M. Hartmann<sup>a</sup>, Emie Durighello<sup>a</sup>, Olivier Pible<sup>a</sup>, Balbina Nogales<sup>b</sup>, Fabrizio Beltrametti<sup>c</sup>, Rafael Bosch<sup>b</sup>, Joseph A. Christie-Oleza<sup>d</sup>, Jean Armengaud<sup>a,\*</sup>

<sup>a</sup> CEA, DSV, IBEB, Lab Biochim System Perturb, Bagnols-sur-Cèze, F-30207, France

<sup>b</sup> Microbiologia, Departament de Biologia, Universitat de les Illes Balears, Palma de Mallorca, Spain

<sup>c</sup> Actygea Srl, Via R. Lepetit 34, 21040 Gerenzano, Italy

<sup>d</sup> School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV47AL, United Kingdom

## ARTICLE INFO

## Article history:

Received 12 February 2014

Received in revised form 11 April 2014

Accepted 12 April 2014

Available online xxxx

## Keywords:

Screening

Non-model organisms

High-throughput proteomics

Next-generation sequencing

Metaproteomics

Proteogenomics

## ABSTRACT

Blue biotechnology, in which aquatic environments provide the inspiration for various products such as food additives, aquaculture, biosensors, green chemistry, bioenergy, and pharmaceuticals, holds enormous promise. Large-scale efforts to sequence aquatic genomes and metagenomes, as well as campaigns to isolate new organisms and culture-based screenings, are helping to push the boundaries of known organisms. Mass spectrometry-based proteomics can complement 16S gene sequencing in the effort to discover new organisms of potential relevance to blue biotechnology by facilitating the rapid screening of microbial isolates and by providing in depth profiles of the proteomes and metaproteomes of marine organisms, both model cultivable isolates and, more recently, exotic non-cultivable species and communities. Proteomics has already contributed to blue biotechnology by identifying aquatic proteins with potential applications to food fermentation, the textile industry, and biomedical drug development. In this review, we discuss historical developments in blue biotechnology, the current limitations to the known marine biosphere, and the ways in which mass spectrometry can expand that knowledge. We further speculate about directions that research in blue biotechnology will take given current and near-future technological advancements in mass spectrometry.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction

Aquatic environments are a tantalizing reservoir of organisms and inestimable source of value for the development of novel biotechnologies. Approximately 70% of the Earth's biosphere is aqueous (oceans, lakes, rivers, ponds, etc.) and oceans represent more than 95% of the water surface (Shiklomanov, 1993). Indeed, oceans are already an industrially relevant source of “farmed” products for a constantly increasing population; there were more than 78 million tons of aquaculture products estimated in 2010 (FAO Fisheries Aquaculture Information Statistics Service, 2012). Aquatic environments are also major sources of new biomedical compounds and products of biotechnological interest (Blunt et al., 2014). However, the vast majority of potentially useful organisms living in these environments remain unexplored, despite massive efforts to sequence marine genomes and metagenomes. Mass spectrometry-based methods can aid in the effort to characterize marine life by facilitating the rapid screening of new isolates. Moreover,

next-generation proteomics promises in depth profiling of the proteomes, exoproteomes, and metaproteomes of marine organisms (Armengaud, 2013). Many of these advances in proteomic approaches are due to the expansion of protein sequence databases, which are growing exponentially thanks to next-generation nucleic acid sequencing and appropriate tools to annotate genomes and metagenomes. Proteomics can be applied to model cultivable isolates, but also exotic or non-cultivable species, and even consortia or communities (Armengaud et al., 2014). Here, we present the potential advantages proteomics can bring to blue biotechnology. This review discusses the use of mass spectrometry to screen novel biotechnology-relevant organisms. We also present how shotgun proteomic approaches could be helpful in understanding the functions of these organisms at the molecular level and in identifying novel biotechnology-relevant enzymes. We further give our perspectives how future studies of aquatic environments will be guided by technological enhancements, specifically in the field of mass spectrometry.

### 2. Blue biotechnology

Blue biotechnology describes the biotechnological applications derived from aquatic environments. The other “color code” biotechnologies

\* Corresponding author at: Laboratoire de Biochimie des Systèmes Perturbés, CEA Marcoule, DSV, IBEB, SBTN, LBSP, F-30207 BAGNOLS-SUR-CEZE, France. Tel.: +33 4 66 79 68 02; fax: +33 4 66 79 19 05.

E-mail address: [jean.armengaud@cea.fr](mailto:jean.armengaud@cea.fr) (J. Armengaud).



are green (agricultural processes), white (industrial processes), red (medical applications), and grey (environmental applications). The boundaries between these fields are hazy because their definitions include both the exploitation of resources and their applications, so a certain degree of overlap exists. Blue biotechnology thus contributes to the other fields, as described below. One of the most notable contributions to arise from blue biotechnology, the Taq polymerase enzyme obtained from the thermophilic bacterium *Thermus aquaticus*, has considerably changed molecular biology and genetic diagnostics. This thermophilic polymerase made the polymerase chain reaction (PCR) that amplifies specific DNA sequences easy to perform, leading to numerous improvements in our understanding of biology and further allowing the development of gene therapy and rapid genetic testing, two pillars of the personalized medicine of tomorrow.

In addition to its contributions to science, blue biotechnology has given rise to many common products, such as agar and a variety of newly discovered exopolysaccharides from marine bacteria used as thickening, coagulating, adhesion, stabilizing, and gelling agents with industrial applications (Satpute et al., 2010). Marine proteins, such as proteases (Reid et al., 1980; Chi et al., 2007; Haddar et al., 2009), lipases (Mo et al., 2009), alginate lyases (Wong et al., 2000; Alkawash et al., 2006; Gacesa, 1992), agarases (Aoki et al., 1990; Belas, 1989; Sugano et al., 1993a, 1993b, 1994), chitinases and chitosanases (Osawa and Koga, 1995), carrageenases (Sarwar et al., 1987; Ohta and Hatada, 2006), cellulose and hemicellulose hydrolases (Tong et al., 1980; Raghukumar et al., 2004; Hou et al., 2006), amylases, fucoidanases, mannanases (Li et al., 2007; Chakraborty et al., 2009; Mohapatra et al., 1998; Furukawa et al., 1992; Yaphe and Morgan, 1959; Yahata et al., 1990; Araki and Kitamikado, 1981), as well as metabolites and biochemicals, have also been listed for food applications. The development of food products and ingredients of marine origin is vital to feed the continuously growing human population (FAO Fisheries and Aquaculture Department, 2012). For example, better bio-inspired treatments could improve the yield, quality, and sustainability of aquaculture, which is subject to infectious diseases and parasites (Oliva-Teles, 2012). Recent developments in shotgun proteomics for non-model organisms will boost research on these blights and reveal biomarkers for monitoring specific organisms (Armengaud et al., 2014).

Marine and freshwater toxicology is a multi-faceted field devoted to the study of toxins from marine organisms and their adverse effects (Schmitt and De Haro, 2013), as well as the characterization of ecological disturbances by pollutants and global warming, i.e., ecotoxicology. The marine symbiont *Vibrio fischeri* is already the reference microorganism for the quantification of pollutants in water (International Standard ISO 11348). The development of aquatic biosensors, including protein biomarkers, to measure *in situ* changes in ecology, climate, and anthropogenic effects is at the interface of biology and microelectronics and promises to be the champion of blue biotechnology (Kroger and Law, 1903–1913; Kroger et al., 2002).

Bioenergy production is a lofty challenge, the full realization of which will likely spill over into future generations. Improving biomass degradation is a first step towards addressing this challenge. Novel degradative enzymes could be extracted from aquatic organisms, as many yet uncharacterized degraders surely exist. Specific culture of algae, microalgae, or other microorganisms able to efficiently produce useful compounds in bulk could be established for use as biofuels (Liu et al., 2013). Various aquatic microorganisms, such as cyanobacteria and algae, are capable of producing hydrogen from solar energy and as such are the subject of intense investigation (Gupta et al., 2013). Microalgal oil production is also considered a viable alternative for biodiesel (Delrue et al., 2013; Gong and Jiang, 2011). While microalgae are considered the most promising group of marine organisms for renewable energy (Shurin et al., 2013), many other organisms could also be of interest for the production of more specific high-value chemical compounds for green chemistry (Hu et al., 2013).

A wide variety of bioactive compounds, such as polysaccharides, lipids, vitamins, sterols, pigments, toxins, alkaloids, and pharmaceuticals, can also be exploited from aquatic organisms (Singh et al., 2005; Tan, 2007; Wijffels et al., 2013). Since the discovery of the first antibiotic, penicillin, in 1929, soil microbiota have been subjected to extensive investigation due to the ease of microbial isolation from the lithosphere compared to any other spheres of earth. The development of high-throughput techniques for pharmacological screening has allowed the identification of a wide variety of novel chemical compounds from microbes. The rate of development of novel drugs has slowed because chemical structures from terrestrial microorganisms are repetitive, but marine microorganisms are often taxonomically unique, which makes them interesting as potential sources of new drugs. Holothurin, an anti-tumor metabolite extracted from the sea cucumber *Actynopygia agassizi* opened the search for new molecules from marine organisms in 1950 (Nigrelli et al., 1967). Since then, new molecules have been identified from sponges (the cytostatic cytarabine, the antiviral compound Ara-A, the anti-tumor compound Ara-C) (Thakur and Muller, 2004; Guyot, 2000), mollusks (e.g., the analgesic ziconotide), bryozoans (e.g., the bryostatin protein kinase C), tunicates, etc. (Haefner, 2003). Pharmaceuticals obtained from marine organisms that have been commercialized for clinical use include ecteinascidin 743, also known as trabectedin (Lichter et al., 1975; Sigel et al., 1979), which has been approved in Europe, Russia, and South Korea for the treatment of soft tissue sarcoma, and ziconotide, which has been approved for intractable pain (Skov et al., 2007). Interest in the exploitation of proteins involved in passive resistance to UV damage or active cellular repair is on the increase, as evidenced by the number of proteomic studies of aquatic organisms able to withstand high levels of UV irradiation (Matallana-Surget et al., 2013, 2012). Furthermore, marine natural products are more likely than their terrestrial counterparts to have antitumor properties (Vinothkumar and Parameswaran, 2013). This tendency reinforces the idea that marine organisms could play a considerable part in this search for new pharmaceuticals (Bhatnagar and Kim, 2010; Debbab et al., 2010), but much of the marine environment is poorly accessible and thus largely unexplored as a resource. Many of the compounds of interest are produced by marine microorganisms, even some originally thought to come from eukaryotes like coral (Wilson et al., 2014). Despite the enthusiasm for marine microorganisms, cultivation techniques for recalcitrant species remain limited and current research is almost exclusively dedicated to genomic approaches. The specific literature on the techniques and culture media for the cultivation of marine microbes is scarce, although some attempts have met with a certain degree of success (Connon and Giovannoni, 2002; Fenical and Jensen, 2006; Rappe et al., 2002; Song et al., 2009). To fully exploit the pharmaceutical potential of biodiversity, adequate technologies are required.

Numerous innovative investigations have been conducted to screen for new organisms from aquatic environmental samples with the goal of discovering new biotechnological products (Imhoff et al., 2011). Their use in molecular biology or in everyday commercial products such as laundry detergent has become common as of a decade ago. Enzymes with novel properties, such as improved stability or better yield (Niehaus et al., 1999; Adams et al., 1995), can be obtained from exotic organisms, such as thermophiles, psychrophiles, and barophiles (Egorova and Antranikian, 2005; van den Burg, 2003). Indeed, bioactive metabolites with wholly unprecedented structures have been described from marine environmental samples. More new marine compounds are described every year as illustrated in Fig. 1 (Blunt et al., 2013). With the development of high-throughput screening methods and novel technologies for their characterization, the number of discovered compounds is now twice the number recorded a decade ago.

### 3. The minuscule portion of known marine biosphere

The marine biosphere is extremely vast and diverse, with some areas among the most hostile environments on Earth, as exemplified by the



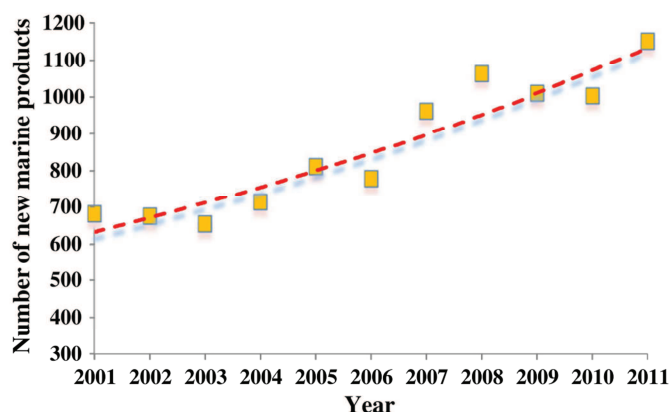


Fig. 1. New marine products described per year as documented by Blunt et al. (2013). A correlation curve shows the increasing trend over the last decade.

Lake Medee, a deep-sea salt-saturated lake in the Eastern Mediterranean sea (Yakimov et al., 2013). The wide range of temperatures (hydrothermal vents, cold seeps, open oceans), salinities (from common seawater to hypersaline brine pools), and oxygen concentrations (from aerobic to anoxic) found in marine environments provide habitats for an extended biodiversity. The full extent of the portion of known organisms in the whole marine biosphere is an open question. The best characterized massive-sampling and metagenome sequencing effort of the ocean surface ever performed is probably the Sorcerer II Global Ocean Sampling. From this initiative, the 16S rDNA-based microbial diversity analysis revealed the presence of more than 800 distinct ribotypes (97% nucleotide identity cutoff), 60 of which accounted for more than 70% of all 16S rDNA sequences (Rusch et al., 2007). The most abundant organisms based on 16S rDNA sequence counts were bacteria: *Alphaproteobacteria* (32–51%), *Gammaproteobacteria* (13–14%), *Bacteroidetes* (8–13%), *Cyanobacteria* (5–8%) and *Actinobacteria* (5–7%) (Rusch et al., 2007; Biers et al., 2009). The distribution of major components of bacterioplankton in those 5 phyla agrees with an earlier report on global diversity (Pommier et al., 2007), as well as with more recent reports (Yilmaz et al., 2012). Despite this general picture, only a few ribotypes belonging to *Alphaproteobacteria* (SAR11-relatives) and *Gammaproteobacteria* (SAR86-relatives) are ubiquitously abundant (Rusch et al., 2007; Biers et al., 2009), thus suggesting that those five cosmopolitan phyla are widely diverse in marine environments. Since then, new massive samplings have been done to further explore the global ocean ecosystem, especially protists and metazoans, as exemplified by the Tara Oceans project (Karsenti et al., 2011). Studies of hypersaline anoxic deep-sea basins show that many barriers to biological dispersal are likely to exist, adding extra evolutionary pressure and thus increasing the diversity of archaeal, bacterial, and eukaryotic microbial life (Stock et al., 2012; Filker et al., 2013; Daffonchio et al., 2006; La Cono et al., 2011; Ferrer et al., 2012).

Currently, the scientific community is engaged in an effort to sequence the whole genomes of marine isolates to investigate the ecological and biotechnological relevance of the wide diversity of marine organisms. For example, nearly 200 genomes of cultivated marine bacteria affiliated with the five cosmopolitan phyla mentioned above were used in recruitment experiments with metagenomic reads from the Sorcerer II Global Ocean Sampling (Yooseph et al., 2010). Surprisingly, less than 25% of all reads recruited against marine genomes, using a relaxed cutoff of 50% of overall nucleotide identity (Yooseph et al., 2010). Moreover, when analysis was restricted to 16S rDNA using 97% of overall nucleotide identity as cutoff (a value widely used for the definition of an Operational Taxonomic Unit (OTU) in environmental surveys), only 33% of 16S rDNA reads recruited against 16S rDNA genes of isolates (Yooseph et al., 2010). These data reinforce the evidence accumulated since the early days of the use of molecular biology

techniques to explore prokaryotic diversity, which indicated that the majority of ocean bacterial diversity remains uncultured. Single cell genomics was successfully proposed as an approach to recover the complete genomes of uncultured bacteria (Stepanaukas, 2012). Sequence information from 56 amplified single genomes from marine organisms was recently compared with available marine metagenomes (Swan et al., 2013). Similar to the previous report regarding genomes of cultured marine bacteria, most marine metagenome reads did not recruit the genomes of uncultured bacteria, i.e., 5% of recruitment using 60% of overall nucleotide identity as a cutoff (Yooseph et al., 2010). Thus, it can be concluded that most of the marine bacterioplankton diversity remains genomically unexplored, and a major effort to obtain information on both cultivable and non-cultivable organisms is needed to clarify the ecological and biotechnological relevance of marine organisms.

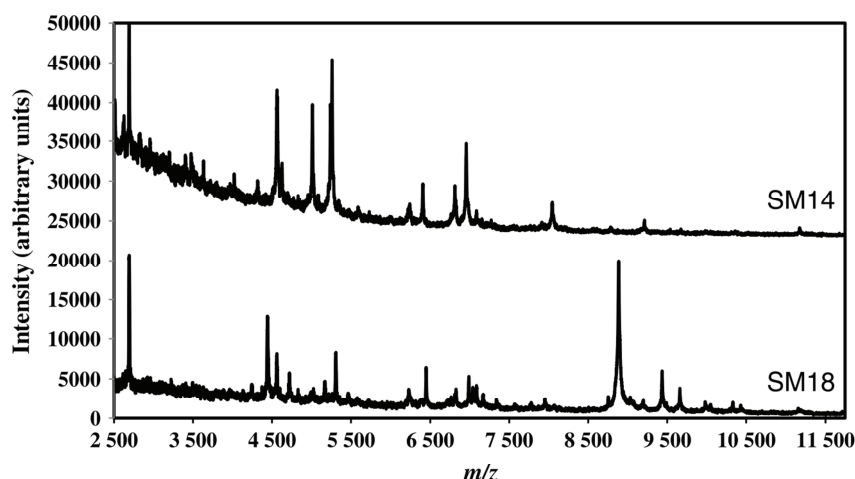
#### 4. Mass spectrometry-based methods for screening new isolates

Mass spectrometry is a widely used technique for the identification and characterization of chemical and biological compounds. Several mass spectrometry detectors, coupled to electrospray or matrix-assisted laser desorption/ionization (MALDI), are currently available: time of flight (TOF), quadrupole, ion trap, orbitrap, or Fourier-transform ion cyclotron resonance. These ionization source and detector combinations cover the majority of instruments used for the characterization of peptides and proteins, although other platforms exist. Over the last decade, mass spectrometry has gained popularity for the identification of microorganisms (Lartigue, 2013; Carbonnelle et al., 2011; Biswas and Rolain, 2013). Mass spectrometry-based methods rely on the analysis of global molecular weights of the abundant, basic, small proteins from the organism to produce an identification via comparison with a database where such profiles have been already recorded. The preferred instrument for this analysis is the MALDI-TOF mass spectrometer. Fundamental and applied research laboratories and hospitals are equipped with such technology for routine bacterial identifications. This mass spectrometer is simple to operate, its operational costs are low, and the acquisition speed (5 min per organism) allows high throughput analysis (Coltella et al., 2013; Fenselau and Demirev, 2001; Kliem and Sauer, 2012; Sauer and Kliem, 2010).

Briefly, the workflow for MALDI-TOF bacterial identification starts with an isolation step where the bacteria are streaked on agar plates followed by overnight culture; bacteria can subsequently be spotted directly on the MALDI-TOF target, a metallic plate. Alternatively, proteins can be extracted using a chemical process prior to spotting (Lasch et al., 2008; Sedo et al., 2011). Spectra for bacterial identification are usually acquired in the 2000–12,000  $m/z$  (mass-to-charge ratio) range (Fig. 2). Observed signals mainly arise from ribosomal bacterial proteins (Ryzhov and Fenselau, 2001). The protein mass pattern is then compared to the profiles obtained from thousands of bacteria, which allows the identification of organisms at the genus, species, and sometimes the subspecies level (Sandrin et al., 2013; Seibold et al., 2010). Commercial suppliers provide the software program for identification together with large libraries of reference spectra; individual users can also create homemade libraries. Because mass pattern-based identification relies on the proteins present in a given physiological condition, it is essential to record the protein profile from bacteria grown and harvested in similar conditions to the reference (Chen et al., 2008; Williams et al., 2003). MALDI-TOF identification and classification are comparable to 16S RNA sequencing (Sauer et al., 2008), and this technique is currently widely implemented in clinical diagnosis (Clark et al., 2013).

While identification using MALDI-TOF works well for most pathogens, it is poorly applicable to new environmental isolates that are distantly related to known organisms. For an accurate identification, the profile of the microorganism of interest has to be present in the database; even closely related species may be insufficient for identification. This problem is not critical in clinical microbiology, where well-defined pathogenic isolates are recurrently found during diagnosis,





**Fig. 2.** Representative whole-cell MALDI-TOF mass spectra of two marine bacteria. The identity of each bacterium can be obtained from the comparison of these spectra to a database of spectra acquired under the same conditions for all known organisms. SM18 and SM14 are two *Krokinobacter* isolates from sea water sampled (2012-07-07) at the beach of Saintes-Maries-de-la-Mer, France (43.4510372N-4.4361055E).

but it becomes a problem for environmental isolates where microbial diversity is overwhelming (Welker and Moore, 2011). One way to overcome this problem is by identifying strong reliable MALDI-TOF biomarkers to screen for a specific group of environmental organisms using shotgun proteomics, such as has been performed for the genus *Ruegeria* (Christie-Oleza et al., 2013a), which is an abundant genus in marine environments. Proteogenomics, wherein proteomic data are used to better predict gene products (Armengaud, 2009; Armengaud et al., 2013), was instrumental in determining the exact molecular weight expected for each biomarker. This concept of specific MALDI-TOF biomarkers has since been extended to the whole *Roseobacter* clade (Christie-Oleza et al., 2013b). The *Roseobacter* lineage is a pertinent group of marine microorganisms due to its abundance in coastal systems and relevance in global biogeochemical cycles (Buchan et al., 2005; Wagner-Dobler and Biebl, 2006). The most interesting trait of this clade is its generalist lifestyle, which confers a large array of encoded functions to scavenge sources of carbon and energy from its diverse marine habitat (Christie-Oleza et al., 2012; Newton et al., 2010). The most reliable MALDI-TOF biomarkers for the clade, in terms of abundance and low variability among strains, were the ribosomal proteins L29 and L30 and the DNA binding protein HU (Christie-Oleza et al., 2013b). These defined biomarkers not only proved effective in the identification of *Roseobacter* strains from a wide variety of natural marine organisms, but also in dereplicating identical isolates. Further work needs to be done to test these biomarkers on a larger range of marine lineages. Nevertheless, whole cell MALDI-TOF MS can be used to identify environmental microorganisms once a set of biomarkers has been defined. The systematic screening of marine isolates by MALDI-TOF represents a tremendous task but marine microbiologists made a start on such challenge.

## 5. Mass spectrometry-based proteomics

Mass spectrometry can be used not only to classify environmental isolates, as discussed above, but also to survey all of the proteins present in a given organism for a given phenotype, i.e., a given proteome. Several reviews on mass spectrometry-based proteomics (Aebersold and Mann, 2003; Domon and Aebersold, 2006; Fenselau, 2007; Hartmann et al., in press), and its application to marine samples in particular (Slattery et al., 2012), are available. In general, proteins can be identified using either peptide mass fingerprinting or tandem mass spectrometry (Fig. 3).

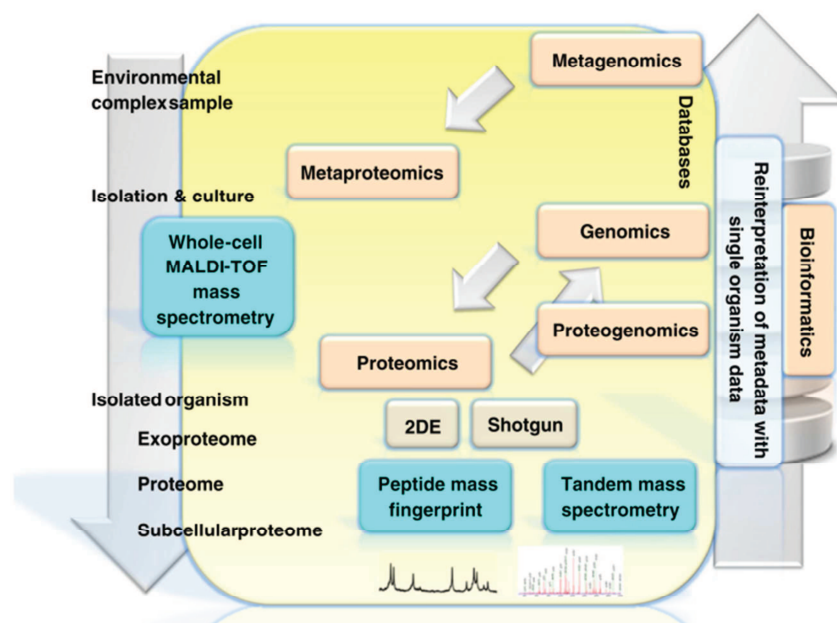
One recent study that took advantage of protein identification using peptide mass fingerprinting is the proteomic investigation of salt tolerance in *Lactobacillus sakei* CRL1756 (Belfiore et al., 2013). This bacterium

is used in the fermentation of anchovy-based food products, where salt concentrations can be lethal to other lactic acid bacteria. To identify the mechanism of salt tolerance used by *L. sakei*, proteins were extracted from cultures grown in control and high-salt conditions. To compare the two phenotypes, proteins were separated using two-dimensional electrophoresis, wherein proteins are spatially distributed across a gel according to their molecular weight and isoelectric point, and excised for mass spectrometry analysis using MALDI-TOF MS (described above). By combining the relative quantitation of gel electrophoresis with identification using mass spectrometry, the authors were able to detect several stress response and metabolic proteins involved in the adaptation of *L. sakei* to high-salt conditions.

In addition to relative quantitation, specific electrophoresis techniques can be used to select proteins with post-translational modifications, such as the phosphorylated proteins detected in a recent study of the aquatic larvae of *Simulium vittatum*, a silk-producing arthropod (Papanicolaou et al., 2013). The phosphorylation sites identified in this study likely confer different physicochemical properties related to the multiple functions of silk in *S. vittatum*. Silks produced by aquatic organisms are of interest because they have unique physicochemical properties making them more resistant to water and shear forces than their terrestrial counterparts. In this study, the major protein components of *S. vittatum* were separated using two-dimensional electrophoresis as in the previous example, but they were then visualized with a stain specific to phosphoproteins. These phosphoproteins were then identified using MALDI-TOF/TOF MS, which is similar to the MALDI-TOF instruments described above but with the addition of a collision cell, where selected peptides are fragmented, and a second time-of-flight detector that records the mass-to-charge ratios of the product ions, yielding more sequence information than peptide mass fingerprinting. The use of two detectors in sequence to record parent and product ion masses is referred to as tandem mass spectrometry.

Many tandem mass spectrometry techniques involve the chromatographic separation of peptides prior to mass spectrometry analysis. As such, many instruments use electrospray ion sources as opposed to MALDI because they can interface directly with the chromatography component. One recent study that made use of liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is the investigation of the mode of action of albamiramide, which are marine-derived antiparasitics, against the human parasite *Trypanosoma brucei* (Sanchez et al., 2013). Here, affinity capture was used to selectively enrich *T. brucei* proteins that interacted with biotinylated or immobilized albamiramides. These enriched proteins were then identified using a high-resolution LC-MS/MS system that performs peptide sequencing,





**Fig. 3.** The main experimental approaches and methodologies in the field of protein mass spectrometry. In peptide mass fingerprinting, spectra are generated from the peptides of partially purified proteins and the observed masses are matched with the theoretical masses of the peptides predicted from known protein sequences. In tandem mass spectrometry, peptides are further fragmented into product ions, the masses of which can then be used to confirm the sequence of the precursor peptide.

where peptide sequences can be deduced by calculating the difference in mass between incrementally longer fragment ions. The identified proteins indicated that albamiramide targets the glycosome, suggesting that these compounds would have low toxicity to higher vertebrates and are thus promising antiparasitic drugs.

Although mass spectrometry-based proteomics has been used with a considerable degree of success on aquatic samples, some subjects may present extra challenges. In addition to the standard difficulties in proteomic sample analysis, marine samples require special treatment regarding handling and transport from the sampling location, salt removal, and data analysis, especially in the case of incomplete genome databases (Slattery et al., 2012). These first two areas of concern should be addressed in the study design and selection of specific sample preparation methods. To address this last concern, it is possible to use proteogenomics, the combination of proteomics and genomics, to obtain meaningful results even in the absence of a fully assembled genome sequence (Armengaud et al., 2013). For example, proteomic data was used to augment shotgun genome sequencing of the aquatic bacterium *Tistlia consotensis*, thereby allowing rapid genome assembly and annotation and a proteomic exploration of the mechanisms of salt tolerance in this organism (Rubiano-Labrador et al., 2014). It is also possible to assess the proteins released outside of the cells (Armengaud et al., 2012), the so-called exoproteome (Fig. 3). Because enzymes released into the environment are more robust than their intracellular counterparts, they are likely to be more relevant to biotechnological applications. As such, exoproteins are attractive targets and are scrutinized with specific proteomic approaches (Durighello et al., 2014). The ability to detect and quantify exoproteins reflects one major advantage of proteomics over genomics or transcriptomics for the identification of relevant aquatic products. Thus despite potential challenges, mass spectrometry-based proteomics can contribute to the discovery and development of blue biotechnologies, as seen in this brief survey of mass spectrometry methods and recent studies that took advantage of them.

## 6. Meta-approaches for global understanding of the biosphere

While proteomics can certainly be useful in certain situations, such as those described above, it is unreasonable to try to culture all marine organisms and reproduce environmental community structures and

phenotypes. Even among cultivable organisms, phenomena such as quorum sensing and predation evasion may be prerequisite to the expression of interesting compounds (e.g., antibiotics). These phenomena can only be observed using environmental meta-approaches, as they may not occur in pure culture. Therefore, environmental meta-approaches, such as metagenomics (Volckaert et al., 2008) and metaproteomics (Wang et al., 2014), wherein genomic and proteomic information are collected from whole communities *in situ*, are essential. Metagenomics gives an exhaustive view of the genetic potential present in a given sample, while metaproteomics gives a view of the active pathways. Mass spectrometry-based proteomics, in combination with metabolomics or mass spectrometric detection of small signaling molecules, is an ideal approach to understanding these situations (Burgess, 2012). While meta-approaches promise a wealth of information, metaproteomic investigations face two major challenges. The first is difficulties related to representative protein extraction, and the second is related to data interpretation in the paucity of genomic data (Fig. 3). Despite the challenges, several marine metaproteomic studies have already been undertaken (see (Wang et al., 2014) for a recent review). Based on a survey of PubMed (performed on April 9, 2014), 17 of the last 100 metaproteomic studies were devoted to aquatic samples.

Marine organisms, especially those in biofilms or adapted to live in extreme environments, such as psychrophiles or barophiles or hyperthermophiles, may be recalcitrant to typical lysis methods, inhibiting protein extraction or modifying the yield of this important step. The hyperthermophile and radiotolerant archaeon *Thermococcus gammatolerans* which was isolated about 2000 m deep in the Pacific Ocean is a typical example of such difficult non-model organism (Zivanovic et al., 2009). Organisms adapted to live in cold, high-pressure environments have specific membrane characteristics to resist shocks in temperature and pressure (Usui et al., 2012), which could render them more difficult to lyse. Individual members of a given community may be preferentially lysed and certain types of proteins preferentially extracted depending on the method used, confounding metaproteomic analyses even when the corresponding metagenomic information is available (Leary et al., 2013, 2012).

Metagenomic sequence data, especially sequences corresponding directly to the metaproteomic sample in question, are almost never available, and currently, identification by homology using sequences



from closely related species is difficult (Armengaud et al., 2014), especially because only a tiny portion of marine organisms have been sequenced. The increase in genomic information will greatly facilitate marine metaproteomic analyses. Furthermore, while single-cell genome sequencing can help in the assembly of metagenomes, reference genomes from cultured organisms are still immensely useful in that they provide the biochemical information needed to properly annotate environmental genomes and thus interpret proteomic data (Rappe, 2013).

## 7. Technology-driven perspectives

Over the last decade, proteomics has evolved considerably from requiring significant effort to resolve proteins prior to costly mass spectrometry analysis on the most interesting protein spots into a rapid and in depth quantitative analysis of thousands of proteins in nearly any organism. This evolution was made possible thanks to i) the avalanche of genome information made accessible by next-generation sequencers, ii) the shotgun concept, which is essentially the identification of all proteins through the characterization of their peptides resulting from proteolysis, and iii) the tremendous evolution of the mass spectrometers used for peptide analysis. Because new generations of mass spectrometers are being built with higher accuracy, increased scan speed, and improved sensitivity, new concepts will emerge to analyze more complex samples. For example, high-resolution and high-speed approaches for the analysis of entire proteins, i.e., top-down approaches, will likely become more popular. Highly multiplexed absolute quantitation will also become increasingly feasible. These novelties will open a wealth of technology-driven perspectives for blue biotechnology. Screening of aquatic and marine organisms will be fast and reliable, paving the way for a wide range of new biotechnological products derived from better knowledge of marine proteins. Knowledge of fundamental molecular mechanisms of the most relevant organisms will increase. Global metaproteomic approaches will also benefit from these technological and conceptual improvements. With successful technology transfer from academia and research institutions to industry, blue biotechnology applications will continue multiply.

## Acknowledgments

We would like to thank Guylaine Miotello for her help recording the MALDI-TOF mass spectra. This work was supported by the Agence Nationale de la Recherche (ANR-12-BSV6-0012-01).

## References

- Adams, M.W., Perler, F.B., Kelly, R.M., 1995. Extremozymes: expanding the limits of biocatalysis. *Biotechnology (NY)* 13, 662–668.
- Aebbersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Alkawas, M.A., Sothill, J.S., Schiller, N.L., 2006. Alginate lyase enhances antibiotic killing of mucoid *Pseudomonas aeruginosa* in biofilms. *APMIS* 114, 131–138.
- Aoki, T., Araki, T., Kitamikado, M., 1990. Purification and characterization of a novel beta-agarase from *Vibrio* sp. AP-2. *Eur. J. Biochem.* 187, 461–465.
- Araki, T., Kitamikado, M., 1981. Beta-mannanase of bacteria isolated from natural habitats. *B. Jpn. Soc. Sci. Fish* 47, 753–760.
- Armengaud, J., 2009. A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin. Microbiol.* 12, 292–300.
- Armengaud, J., 2013. Microbiology and proteomics, getting the best of both worlds! *Environ. Microbiol.* 15, 12–23.
- Armengaud, J., Christie-Oleza, J.A., Clair, G., Malard, V., Duport, C., 2012. Exoproteomics: exploring the world around biological systems. *Expert Rev. Proteomics* 9, 561–575.
- Armengaud, J., Hartmann, E.M., Bland, C., 2013. Proteogenomics for environmental microbiology. *Proteomics* 13, 2731–2742.
- Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M., 2014. Non-model organisms, a species endangered by proteogenomics. *J. Proteomics*.
- Belas, R., 1989. Sequence analysis of the *agrA* gene encoding beta-agarase from *Pseudomonas atlantica*. *J. Bacteriol.* 171, 602–605.
- Belfiore, C., Fadda, S., Raya, R., Vignolo, G., 2013. Molecular basis of the adaption of the anchovy isolate *Lactobacillus sakei* CRL1756 to salted environments through a proteomic approach. *Food Res. Intern.* 54, 1334–1341.
- Bhatnagar, I., Kim, S.K., 2010. Immense essence of excellence: marine microbial bioactive compounds. *Mar. Drugs* 8, 2673–2701.
- Biers, E.J., Sun, S., Howard, E.C., 2009. Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl. Environ. Microbiol.* 75, 2221–2229.
- Biswas, S., Rolain, J.M., 2013. Use of MALDI-TOF mass spectrometry for identification of bacteria that are difficult to culture. *J. Microbiol. Methods* 92, 14–24.
- Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H., Prinsep, M.R., 2013. Marine natural products. *Nat. Prod. Rep.* 30, 237–323.
- Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H., Prinsep, M.R., 2014. Marine natural products. *Nat. Prod. Rep.* 31, 160–258.
- Buchan, A., Gonzalez, J.M., Moran, M.A., 2005. Overview of the marine roseobacter lineage. *Appl. Environ. Microbiol.* 71, 5665–5677.
- Burgess, J.G., 2012. New and emerging analytical techniques for marine biotechnology. *Curr. Opin. Biotechnol.* 23, 29–33.
- Carbonnelle, E., Mesquita, C., Bille, E., Day, N., Dauphin, B., Beretti, J.L., Ferroni, A., Gutmann, L., Nassif, X., 2011. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clin. Biochem.* 44, 104–109.
- Chakraborty, S., Khopade, A., Kokare, K., Mahadik, K., Chopade, B., 2009. Isolation and characterization of novel alpha-amylase from marine *Streptomyces* sp D1. *J. Mol. Catal. B-Enzym.* 58, 17–23.
- Chen, P., Lu, Y., Harrington, P.B., 2008. Biomarker profiling and reproducibility study of MALDI-MS measurements of *Escherichia coli* by analysis of variance-principal component analysis. *Anal. Chem.* 80, 1474–1481.
- Chi, Z., Ma, C., Wang, P., Li, H.F., 2007. Optimization of medium and cultivation conditions for alkaline protease production by the marine yeast *Aureobasidium pullulans*. *Bioresour. Technol.* 98, 534–538.
- Christie-Oleza, J.A., Fernandez, B., Nogales, B., Bosch, R., Armengaud, J., 2012. Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. *ISME J.* 6, 124–135.
- Christie-Oleza, J.A., Pina-Villalonga, J.M., Guerin, P., Miotello, G., Bosch, R., Nogales, B., Armengaud, J., 2013a. Shotgun nanoLC-MS/MS proteogenomics to document MALDI-TOF biomarkers for screening new members of the *Ruegeria* genus. *Environ. Microbiol.* 15, 133–147.
- Christie-Oleza, J.A., Miotello, G., Armengaud, J., 2013b. Proteogenomic definition of biomarkers for the large roseobacter clade and application for a quick screening of new environmental isolates. *J. Proteome Res.* 12, 5331–5339.
- Clark, A.E., Kaleta, E.J., Arora, A., Wolk, D.M., 2013. Matrix-assisted laser desorption/ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* 26, 547–603.
- Coltella, L., Mancinelli, L., Onori, M., Lucignano, B., Menichella, D., Sorge, R., Raponi, M., Mancini, R., Russo, C., 2013. Advancement in the routine identification of anaerobic bacteria by MALDI-TOF mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.* 32, 1183–1192.
- Connon, S.A., Giovannoni, S.J., 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68, 3878–3885.
- Daffonchio, D., Borin, S., Brusa, T., Brusetti, L., van der Wielen, P.W., Bolhuis, H., Yakimov, M.M., D'Auria, G., Giuliano, L., Marty, D., Tamburini, C., McGenity, T.J., Hallsworth, J.E., Sass, A.M., Timmis, K.N., Tselepidis, A., de Lange, G.J., Hubner, A., Thomson, J., Varnavas, S.P., Gasparoni, F., Gerber, H.W., Malinverno, E., Corselli, C., Garcin, J., McKew, B., Golyshin, P.N., Lampadariou, N., Polymenakou, P., Calore, D., Cenedese, S., Zanon, F., Hoog, S., 2006. Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. *Nature* 440, 203–207.
- Debbab, A., Aly, A.H., Lin, W.H., Proksch, P., 2010. Bioactive compounds from marine bacteria and fungi. *Microb. Biotechnol.* 3, 544–563.
- Delrue, F., Li-Beisson, Y., Setier, P.A., Sahut, C., Roubaud, A., Froment, A.K., Peltier, G., 2013. Comparison of various microalgae liquid biofuel production pathways based on energetic, economic and environmental criteria. *Bioresour. Technol.* 136, 205–212.
- Domon, B., Aebersold, R., 2006. Mass spectrometry and protein analysis. *Science* 312, 212–217.
- Durighello, E., Christie-Oleza, J.A., Armengaud, J., 2014. Assessing the exoproteome of marine bacteria, lesson from a RTX-toxin abundantly secreted by *Phaeobacter* strain DSM 17395. *PLoS One* 9, e89691.
- Egorova, K., Antranikian, G., 2005. Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.* 8, 649–655.
- FAO Fisheries and Aquaculture Department, 2012. The State of World Fisheries and Aquaculture 2012 (Rome).
- FAO Fisheries Aquaculture Information Statistics Service, 2012. FAO yearbook. In: Statistics and Information Branch of the Fisheries and Aquaculture Department (Ed.), Fishery and Aquaculture Statistics. 2010 (Rome).
- Fenical, W., Jensen, P.R., 2006. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* 2, 666–673.
- Fenselau, C., 2007. A review of quantitative methods for proteomic studies. *J. Chromatogr. B. Anal. Technol. Biomed. Life Sci.* 855, 14–20.
- Fenselau, C., Demirev, P.A., 2001. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* 20, 157–171.
- Ferrer, M., Werner, J., Chernikova, T.N., Bargiela, R., Fernandez, L., La Cono, V., Waldmann, J., Teeling, H., Golyshina, O.V., Glockner, F.O., Yakimov, M.M., Golyshin, P.N., 2012. Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study. *Environ. Microbiol.* 14, 268–281.
- Filker, S., Stock, A., Breiner, H.W., Edgcomb, V., Orsi, W., Yakimov, M.M., Stoeck, T., 2013. Environmental selection of protistan plankton communities in hypersaline anoxic deep-sea basins, Eastern Mediterranean Sea. *Microbiologyopen* 2, 54–63.
- Furukawa, S.-i., Fujikawa, T., Koga, D., Ioe, A., 1992. Purification and some properties of exo-type fucoidanases from *Vibrio* sp. N-5. *Biosci. Biotechnol. Biochem.* 56, 1829–1834.



- Gacesa, P., 1992. Enzymic degradation of alginates. *Int. J. Biochem.* 24, 545–552.
- Gong, Y., Jiang, M., 2011. Biodiesel production with microalgae as feedstock: from strains to biodiesel. *Biotechnol. Lett.* 33, 1269–1284.
- Gupta, S.K., Kumari, S., Reddy, K., Bux, F., 2013. Trends in biohydrogen production: major challenges and state-of-the-art developments. *Environ. Technol.* 34, 1653–1670.
- Guyot, M., 2000. Intricate aspects of sponge chemistry. *Zoosystema* 22, 419–431.
- Haddar, A., Agrebi, R., Bougateg, A., Hmidet, N., Sellami-Kamoun, A., Nasri, M., 2009. Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. *Bioresour. Technol.* 100, 3366–3373.
- Haefner, B., 2003. Drugs from the deep: marine natural products as drug candidates. *Drug Discov. Today* 8, 536–544.
- Hartmann, E.M., Allain, F., Gaillard, J.C., Pible, O., Armengaud, J., 2014. Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol. Biol.* (in press).
- Hou, Y.H., Wang, T.H., Long, H., Zhu, H.Y., 2006. Novel cold-adaptive *Penicillium* strain F5010 secreting thermo-labile xylanase isolated from Yellow Sea. *Acta Biochim. Biophys. Sin. (Shanghai)* 38, 142–149.
- Hu, G., Ji, S., Yu, Y., Wang, S., Zhou, G., Li, F., 2013. Organisms for biofuel production: natural bioresources and methodologies for improving their biosynthetic potentials. *Adv. Biochem. Eng. Biotechnol.* (in press).
- Imhoff, J.F., Labes, A., Wiese, J., 2011. Bio-mining the microbial treasures of the ocean: new natural products. *Biotechnol. Adv.* 29, 468–482.
- Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., Sullivan, M., Arendt, D., Benzioni, F., Claverie, J.M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., Pesant, S., Reynaud, E.G., Sardet, C., Sieracki, M.E., Speich, S., Velayoudon, D., Weissenbach, J., Wincker, P., 2011. A holistic approach to marine eco-systems biology. *PLoS Biol.* 9, e1001177.
- Kliem, M., Sauer, S., 2012. The essence on mass spectrometry based microbial diagnostics. *Curr. Opin. Microbiol.* 15, 397–402.
- Kroger, S., Law, R.J., 1903–1913. Biosensors for marine applications. We all need the sea, but does the sea need biosensors? *Biosens. Bioelectron.* 20 (2005).
- Kroger, S., Piletsky, S., Turner, A.P., 2002. Biosensors for marine pollution research, monitoring and control. *Mar. Pollut. Bull.* 45, 24–34.
- La Cono, V., Smedile, F., Bortoluzzi, G., Arcadi, E., Maimone, G., Messina, E., Borghini, M., Oliveri, E., Mazzola, S., L'Haridon, S., Toffin, L., Genovese, L., Ferrer, M., Giuliano, L., Golyshtin, P.N., Yakimov, M.M., 2011. Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: prokaryotes and environmental settings. *Environ. Microbiol.* 13, 2250–2268.
- Lartigue, M.F., 2013. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry for bacterial strain characterization. *Infect. Genet. Evol.* 13, 230–235.
- Lasch, P., Nattermann, H., Erhard, M., Stammler, M., Grunow, R., Bannert, N., Appel, B., Naumann, D., 2008. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Anal. Chem.* 80, 2026–2034.
- Leary, D.H., Hervey, W.J., Li, R.W., Deschamps, J.R., Kusterbeck, A.W., Vora, G.J., 2012. Method development for metaproteomic analyses of marine biofilms. *Anal. Chem.* 84, 4006–4013.
- Leary, D.H., Hervey, W.J., Deschamps, J.R., Kusterbeck, A.W., Vora, G.J., 2013. Which metaproteome? The impact of protein extraction bias on metaproteomic analyses. *Mol. Cell. Proteom.* 12, 193–199.
- Li, H., Chi, Z., Wang, X., Ma, C., 2007. Amylase production by the marine yeast *Aureobasidium pullulans* N13d. *J. Ocean Univ. China* 6, 60–65.
- Lichter, W., Lopez, D.M., Wellham, L.L., Sigel, M.M., 1975. *Ecteinascidia turbinata* extracts inhibit DNA synthesis in lymphocytes after mitogenic stimulation by lectins. *Proc. Soc. Exp. Biol. Med.* 150, 475–478.
- Liu, H., Cheng, T., Xian, M., Cao, Y., Fang, F., Zou, H., 2014. Fatty acid from the renewable sources: a promising feedstock for the production of biofuels and biobased chemicals. *Biotechnol. Adv.* 32, 382–389.
- Matallana-Surget, S., Villette, C., Intertaglia, L., Joux, F., Bourrain, M., Lebaron, P., 2012. Response to UVB radiation and oxidative stress of marine bacteria isolated from South Pacific Ocean and Mediterranean Sea. *J. Photochem. Photobiol. B* 117, 254–261.
- Matallana-Surget, S., Cavicchioli, R., Fauconnier, C., Wattiez, R., Leroy, B., Joux, F., Raftery, M.J., Lebaron, P., 2013. Shotgun redox proteomics: identification and quantitation of carbonylated proteins in the UVB-resistant marine bacterium, *Photobacterium angustum* S14. *PLoS One* 8, e68112.
- Mo, S., Kim, J.H., Cho, K.W., 2009. Enzymatic properties of an extracellular phospholipase C purified from a marine streptomycete. *Biosci. Biotechnol. Biochem.* 73, 2136–2137.
- Mohapatra, B., Banerjee, U., Bapuji, M., 1998. Characterization of a fungal amylase from *Mucor* sp. associated with the marine sponge *Spirastrella* sp. *J. Biotechnol.* 60, 113–117.
- Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E., Howard, E.C., King, E., Oakley, C.A., Reisch, C.R., Rinta-Kanto, J.M., Sharma, S., Sun, S., Varaljay, V., Vila-Costa, M., Westrich, J.R., Moran, M.A., 2010. Genome characteristics of a generalist marine bacterial lineage. *ISME J.* 4, 784–798.
- Niehaus, F., Bertoldo, C., Kahler, M., Antranikian, G., 1999. Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.* 51, 711–729.
- Nigrelli, R.F., Stempieri Jr., M.F., Ruggieri, G.D., Liguori, V.R., Cecil, J.T., 1967. Substances of potential biomedical importance from marine organisms. *Fed. Proc.* 26, 1197–1205.
- Ohta, Y., Hatada, Y., 2006. A novel enzyme, lambda-carrageenase, isolated from a deep-sea bacterium. *J. Biochem.* 140, 475–481.
- Oliva-Teles, A., 2012. Nutrition and health of aquaculture fish. *J. Fish Dis.* 35, 83–108.
- Osawa, R., Koga, T., 1995. An investigation of aquatic bacteria capable of utilizing chitin as the sole source of nutrients. *Lett. Appl. Microbiol.* 21, 288–291.
- Papanicolaou, A., Woo, A., Brei, B., Ma, D., Masedunskas, A., Gray, E., Guishan Xiao, G., Cho, S., Brockhouse, C., 2013. Novel aquatic silk genes *Simulium* (Psilozia) vittatum (Zett) Diptera: Simuliidae. *Insect Biochem. Mol. Biol.* 43, 1181–1188.
- Pommier, T., Canback, B., Riemann, L., Bostrom, K.H., Simu, K., Lundberg, P., Tunlid, A., Hagstrom, A., 2007. Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* 16, 867–880.
- Raghukumar, C., Muraleedharan, U., Gaud, V.R., Mishra, R., 2004. Xylanases of marine fungi of potential use for biobleaching of paper pulp. *J. Ind. Microbiol. Biotechnol.* 31, 433–441.
- Rappe, M.S., 2013. Stabilizing the foundation of the house that 'omics builds: the evolving value of cultured isolates to marine microbiology. *Curr. Opin. Microbiol.* 16, 618–624.
- Rappe, M.S., Connon, S.A., Vergin, K.L., Giovannoni, S.J., 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633.
- Reid, G.C., Woods, D.R., Robb, F.T., 1980. Peptone induction and rifampin-insensitive collagenase production by *Vibrio alginolyticus*. *J. Bacteriol.* 142, 447–454.
- Rubiano-Labrador, C., Bland, C., Miotello, G., Guerin, P., Pible, O., Baena, S., Armengaud, J., 2014. Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring. *J. Proteomics* 97, 36–47.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkuch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.L., Souza, V., Bonilla-Rosco, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealon, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5, e77.
- Ryzhov, V., Fenselau, C., 2001. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Anal. Chem.* 73, 746–750.
- Sanchez, L.M., Knudsen, G.M., Helbig, C., De Muylder, G., Mascuch, S.M., Mackey, Z.B., Gerwick, L., Clayton, C., McKerrrow, J.H., Linington, R.G., 2013. Examination of the mode of action of the almiramide family of natural products against the kinetoplastid parasite *Trypanosoma brucei*. *J. Nat. Prod.* 76, 630–641.
- Sandrin, T.R., Goldstein, J.E., Schumaker, S., 2013. MALDI TOF MS profiling of bacteria at the strain level: a review. *Mass Spectrom. Rev.* 32, 188–217.
- Sarwar, G., Matayoshi, S., Oda, H., 1987. Purification of a kappa-carrageenase from marine *Cytophaga* species. *Microbiol. Immunol.* 31, 869–877.
- Satpute, S.K., Banat, I.M., Dhakephalkar, P.K., Banpurkar, A.G., Chopade, B.A., 2010. Biosurfactants, bioemulsifiers and exopolysaccharides from marine microorganisms. *Biotechnol. Adv.* 28, 436–450.
- Sauer, S., Kliem, M., 2010. Mass spectrometry tools for the classification and identification of bacteria. *Nat. Rev. Microbiol.* 8, 74–82.
- Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M., Geider, K., 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One* 3, e2843.
- Schmitt, C., De Haro, L., 2013. Clinical marine toxicology: a European perspective for clinical toxicologists and poison centers. *Toxins (Basel)* 5, 1343–1352.
- Sedo, O., Sedlacek, I., Zdrahal, Z., 2011. Sample preparation methods for MALDI-MS profiling of bacteria. *Mass Spectrom. Rev.* 30, 417–434.
- Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., Speltztoesser, W., 2010. Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J. Clin. Microbiol.* 48, 1061–1069.
- Shiklomanov, I., 1993. World fresh water resources. In: Gleick, P.H. (Ed.), *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press, New York, NY.
- Shurin, J.B., Abbott, R.L., Deal, M.S., Kwan, G.T., Litchman, E., McBride, R.C., Mandal, S., Smith, V.H., 2013. Industrial-strength ecology: trade-offs and opportunities in algal biofuel production. *Ecol. Lett.* 16, 1393–1404.
- Sigel, M.M., Lichter, W., Ghaffar, A., Wellham, L.L., Weinheimer, A.J., 1979. Substances from marine organisms influencing tumor growth and immune responses. *Adv. Exp. Med. Biol.* 121, 577–588.
- Singh, S., Kate, B.N., Banerjee, U.C., 2005. Bioactive compounds from cyanobacteria and microalgae: an overview. *Crit. Rev. Biotechnol.* 25, 73–95.
- Skov, M.J., Beck, J.C., de Kater, A.W., Shopp, G.M., 2007. Nonclinical safety of ziconotide: an intrathecal analgesic of a new pharmaceutical class. *Int. J. Toxicol.* 26, 411–421.
- Slattery, M., Ankisetty, S., Corrales, J., Marsh-Hunkin, K.E., Cochfeld, D.J., Willett, K.L., Rimoldi, J.M., 2012. Marine proteomics: a critical assessment of an emerging technology. *J. Nat. Prod.* 75, 1833–1877.
- Song, J., Oh, H.M., Cho, J.C., 2009. Improved culturability of SAR11 strains in dilution-to-extinction culturing from the East Sea, West Pacific Ocean. *FEMS Microbiol. Lett.* 295, 141–147.
- Stepanaukas, R., 2012. Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* 15, 613–620.
- Stock, A., Breiner, H.W., Pachadaki, M., Edgcomb, V., Filker, S., La Cono, V., Yakimov, M.M., Stoeck, T., 2012. Microbial eukaryote life in the new hypersaline deep-sea basin Thetis. *Extremophiles* 16, 21–34.
- Sugano, Y., Terada, I., Arita, M., Noma, M., Matsumoto, T., 1993a. Purification and characterization of a new agarase from a marine bacterium, *Vibrio* sp. strain JT0107. *Appl. Environ. Microbiol.* 59, 1549–1554.
- Sugano, Y., Matsumoto, T., Kodama, H., Noma, M., 1993b. Cloning and sequencing of agaA, a unique agarase 0107 gene from a marine bacterium, *Vibrio* sp. strain JT0107. *Appl. Environ. Microbiol.* 59, 3750–3756.
- Sugano, Y., Matsumoto, T., Noma, M., 1994. Sequence analysis of the agaB gene encoding a new beta-agarase from *Vibrio* sp. strain JT0107. *Biochim. Biophys. Acta* 1218, 105–108.
- Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., Gonzalez, J.M., Luo, H., Wright, J.J., Landry, Z.C., Hanson, N.W., Thompson, B.P., Poulton, N.J., Schwientek, P., Acinas, S.G., Giovannoni, S.J., Moran, M.A., Hallam, S.J., Cavicchioli, R., Woyke, T., Stepanaukas, R., 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11463–11468.



- Tan, L.T., 2007. Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* 68, 954–979.
- Thakur, N.L., Muller, W.E.G., 2004. Biotechnological potential of marine sponges. *Curr Sci India* 86, 1506–1512.
- Tong, C.C., Cole, A.L., Shepherd, M.G., 1980. Purification and properties of the cellulases from the thermophilic fungus *Thermoascus aurantiacus*. *Biochem. J.* 191, 83–94.
- Usui, K., Hiraki, T., Kawamoto, J., Kurihara, T., Nogi, Y., Kato, C., Abe, F., 2012. Eicosapentaenoic acid plays a role in stabilizing dynamic membrane structure in the deep-sea piezophile *Shewanella violacea*: a study employing high-pressure time-resolved fluorescence anisotropy measurement. *Biochim. Biophys. Acta* 1818, 574–583.
- van den Burg, B., 2003. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* 6, 213–218.
- Vinothkumar, S., Parameswaran, P.S., 2013. Recent advances in marine drug research. *Biotechnol. Adv.* 31, 1826–1845.
- Volckaert, F.A., Barbier, M., Canario, A.V., Olsen, J.L., Wesnigk, J., Clark, M., Boyen, C., 2008. Empowering marine science through genomics. *Mar. Genomics* 1, 33–35.
- Wagner-Dobler, I., Biehl, H., 2006. Environmental biology of the marine Roseobacter lineage. *Annu. Rev. Microbiol.* 60, 255–280.
- Wang, D.Z., Xie, Z.X., Zhang, S.F., 2014. Marine metaproteomics: current status and future directions. *J. Proteomics* 97, 27–35.
- Welker, M., Moore, E.R., 2011. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.* 34, 2–11.
- Wijffels, R.H., Kruse, O., Hellingwerf, K.J., 2013. Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Curr. Opin. Biotechnol.* 24, 405–413.
- Williams, T.L., Andrzejewski, D., Lay, J.O., Musser, S.M., 2003. Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. *J. Am. Soc. Mass Spectrom.* 14, 342–351.
- Wilson, M.C., Mori, T., Ruckert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crusemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J., 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506, 58–62.
- Wong, T.Y., Preston, L.A., Schiller, N.L., 2000. ALGINATE LYASE: review of major sources and enzyme characteristics, structure–function analysis, biological roles, and applications. *Annu. Rev. Microbiol.* 54, 289–340.
- Yahata, N., Watanabe, T., Nakamura, Y., Yamamoto, Y., Kamimiya, S., Tanaka, H., 1990. Structure of the gene encoding beta-1,3-glucanase A1 of *Bacillus circulans* WL-12. *Gene* 86, 113–117.
- Yakimov, M.M., La Cono, V., Slepak, V.Z., La Spada, G., Arcadi, E., Messina, E., Borghini, M., Monticelli, L.S., Rojo, D., Barbas, C., Golyshina, O.V., Ferrer, M., Golyshin, P.N., Giuliano, L., 2013. Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci. Rep.* 3, 3554.
- Yaphe, W., Morgan, K., 1959. Enzymic hydrolysis of fucoidin by *Pseudomonas atlantica* and *Pseudomonas carrageenovora*. *Nature* 183, 761–762.
- Yilmaz, P., Iversen, M.H., Hankeln, W., Kottmann, R., Quast, C., Glockner, F.O., 2012. Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiol. Ecol.* 81, 373–385.
- Yooseph, S., Nealson, K.H., Rusch, D.B., McCrow, J.P., Dupont, C.L., Kim, M., Johnson, J., Montgomery, R., Ferriera, S., Beeson, K., Williamson, S.J., Tovchigrechko, A., Allen, A. E., Zeigler, L.A., Sutton, G., Eisenstadt, E., Rogers, Y.H., Friedman, R., Frazier, M., Venter, J.C., 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468, 60–66.
- Zivanovic, Y., Armengaud, J., Lagorce, A., Leplat, C., Guerin, P., Dutertre, M., Anthouard, V., Forterre, P., Wincker, P., Confalonieri, F., 2009. Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biol.* 10, R70.



### 3 Liste des communications réalisées durant cette thèse :

#### 3.1 Présentations orales :

« **Apport de la spectrométrie à haute résolution pour l'identification des composants de l'exoprotéome de bactéries marines** »

**E. Durighello**, J. Christie-Oleza, J. Armengaud.

29<sup>èmes</sup> Journées Françaises de Spectrométrie de Masse – 17-20 septembre 2012 – Orléans.

« **Détection de pathogènes de la menace par spectrométrie de masse** »

**E. Durighello**, A. Lorphelin, M-A. Roncato, L. Bellanger, E. Ezan, J. Armengaud.

Journées Scientifiques de Marcoule – 10-13 juin 2014 – Marcoule.

« **A proteogenomic strategy for defining biomarkers for quick identification of *Francisella* subspecies by MALDI-TOF MS** »

Emie Durighello, Eric Ezan, Laurent Bellanger, **Jean Armengaud**.

The association for Mass Spectrometry: Application to the Clinical Lab (MSACL) 1st annual congress – 2-5 septembre 2014 – Salzburg, Autriche.

#### 3.2 Posters :

« **Shotgun exploration of the protein space : may the MS/MS force be with you, young padawans!** »

**O. Pible**, F. Allain, C. Darolles, V. Malard, P. Guérin, E. Durighello, J. Christie-Oleza, J-C. Gaillard, A. Dedieu, G. Clair, J. Armengaud.

Journée scientifique de l'IBEB – 17 mai 2011 – Marcoule.

« **Identification de microorganismes par spectrométrie : avec ou sans a priori, tous les goûts sont permis !** »

**E. Durighello**, F. Allain, P. Guérin, G. Miotello, A. Lorphelin, L. Bellanger, E. Ezan, O. Pible, J. Armengaud.

Journée scientifique de l'IBEB – 05 avril 2012 – Marseille.

« **Discriminating pathogenic and non-pathogenic *Francisella* strains with three proteogenomic biomarkers** »

E. Durighello, A. Lorphelin, M-A. Roncato, L. Bellanger, **E. Ezan**, J. Armengaud.

61<sup>st</sup> American Society of Mass Spectrometry Conference – 9-13 juin 2013 – Minneapolis MN.

« ***Francisella tularensis* detection in complex media using magnetic immunocapture enrichment and mass spectrometry** »

**E. Durighello**, A. Lorphelin, M-A. Roncato, L. Bellanger, E. Ezan, J. Armengaud.

7<sup>th</sup> European Proteomics Association (EuPA) Conference – 14-17 octobre 2013 – St-Malo.

**« Identification and characterization of *Francisella* strains by MALDI-TOF detection for biodefense purposes »**

**E. Durighello**, A. Lorphelin, M-A. Roncato, L. Bellanger, E. Ezan, J. Armengaud.

62<sup>nd</sup> American Society of Mass Spectrometry Conference – 15-19 juin 2014 – Baltimore, MD.